

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 16:31:11 ; Search time 141.3 Seconds
(without alignments)
7375.802 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878
Sequence: 1 atgagttctcagagctgctgc.....ttcacatgaagatgatga 1878

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	87.8	4.7	7218	1 US-08-232-463-14	Sequence 14, Appl
2	75	4.0	272	4 US-09-313-294A-121	Sequence 121, App
3	74.4	4.0	274	4 US-09-313-294A-463	Sequence 463, App
4	58.8	3.1	3489	2 US-08-728-323A-1	Sequence 1, Appl
5	58.8	3.1	3489	4 US-09-298-568-1	Sequence 1, Appl
6	58.8	3.1	3489	4 US-09-410-399-1	Sequence 1, Appl
7	58.8	3.1	32207	2 US-08-770-379-20	Sequence 20, Appl
8	58.8	3.1	32207	3 US-08-757-669A-20	Sequence 20, Appl
9	58.8	3.1	32207	4 US-09-230-371A-20	Sequence 20, Appl
C 10	53.2	2.8	16442	3 US-08-781-891-208	Sequence 208, App
C 11	53.2	2.8	16442	4 US-09-618-166-208	Sequence 208, App
C 12	49.8	2.7	51259	3 US-08-781-891-209	Sequence 209, App
C 13	49.8	2.7	51259	4 US-09-618-166-209	Sequence 209, App
14	47.4	2.5	1276	3 US-09-411-812A-2	Sequence 2, Appl
15	47.4	2.5	1276	3 US-09-411-812A-2	Sequence 2, Appl
16	47.4	2.5	1276	4 US-09-590-113-2	Sequence 2, Appl
17	47	2.5	6617	4 US-09-976-594-268	Sequence 268, App
18	46.6	2.5	2338	1 US-08-425-069-1	Sequence 1, Appl
19	46.6	2.5	2338	2 US-08-317-844B-1	Sequence 1, Appl
20	45.4	2.4	340	5 US-08-182-175A-104	Sequence 104, App
21	45.4	2.4	340	5 PCT-US92-06412-104	Sequence 104, App
22	43	2.3	1926	4 US-09-249-585A-2	Sequence 2, Appl
23	43	2.3	1926	4 US-09-410-399-3	Sequence 2, Appl
24	43	2.3	2580	4 US-09-050-863-2	Sequence 2, Appl
25	43	2.3	2580	4 US-09-359-081-2	Sequence 2, Appl
C 26	43	2.3	5452	2 US-09-130-114-1	Sequence 1, Appl
C 27	43	2.3	8705	4 US-09-647-344A-14	Sequence 14, Appl

28	43	2.3	9600	3 US-08-910-647-1	Sequence 1, Appl
29	43	2.3	9600	4 US-09-620-925-1	Sequence 1, Appl
30	43	2.3	10596	1 US-07-884-811-15	Sequence 15, Appl
31	43	2.3	10596	1 US-07-885-971-15	Sequence 15, Appl
32	43	2.3	10596	1 US-08-087-783A-15	Sequence 15, Appl
33	43	2.3	10596	1 US-08-194-088B-15	Sequence 15, Appl
34	43	2.3	10596	2 US-08-194-087-15	Sequence 15, Appl
35	43	2.3	10596	5 PCT-US93-04648-15	Sequence 15, Appl
C 36	43	2.3	16080	4 US-09-724-566A-48	Sequence 48, Appl
37	42.2	2.2	1251	4 US-09-828-000-1	Sequence 1, Appl
C 38	42.2	2.2	1270	4 US-08-387-805-1	Sequence 1, Appl
39	42.2	2.2	1558	4 US-09-702-327-3	Sequence 3, Appl
40	41.2	2.2	2277	1 US-08-676-967-2	Sequence 2, Appl
41	41.2	2.2	2277	1 US-08-676-974-2	Sequence 2, Appl
42	41.2	2.2	2277	2 US-09-098-487-2	Sequence 2, Appl
43	41.2	2.2	8920	2 US-08-446-855A-1	Sequence 1, Appl
44	41.2	2.2	8920	3 US-09-150-741-1	Sequence 1, Appl
45	41	2.2	966	2 US-08-766-738-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: PTZ9PT-F18
; US-08-232-463-14
Query Match 4.7%; Score 87.8; DB 1; Length 7218;

Patent No. 5948676
GENERAL INFORMATION:
APPLICANT: Chang, Yuan
APPLICANT: Bohenzky, Roy A.
APPLICANT: Russo, James J.
APPLICANT: Edelman, Isidore S.
APPLICANT: Moore, Patrick S.
TITLE OF INVENTION: Immediate Early Protein From Kaposi's
TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA

TITLE OF INVENTION: Encoding Same And Uses Thereof
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSK/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 3.1%; Score 58.8; DB 2; Length 3489;
Best Local Similarity 43.4%; Pred. No. 6.7e-07;
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

QY 1218 GAAGCAAGATCTGGACATATTCATCACTCTCAAGGCAAAACAAGGCTGAATTGCA 1277
DB 1800 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1859
QY 1278 GTTGAATCATACCAAGAGATGTTGTAAGAGAGCTGAGCAGATCTGAGAGCAATCA 1337
DB 1860 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1919
QY 1338 GCAGCTGAATCTTTAAGAACAGCTCTCAAAAACAAGCAGCAGCAAGGCTGTTGA 1397
DB 1920 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1979
QY 1398 GGAATCTCTGGAATTTATGAGCAGAGAGCTGCTAGAACACTGCAAGAGATTAATCGATCGT 1457
DB 1980 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2039
QY 1458 GAGACAGAGAACTAAGATGACATGCAAGAACAGAGGAGAGATGAGCAGCAG 1517
DB 2040 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2099
QY 1518 GTTTTCATGATTCATCAATCAACAGATCCATGAAAGAGAGCAGCAAGAGAGATTT 1577
DB 2100 GAGACAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2159
QY 1578 CGAGATGTTGACAGCAGCAGAACTGTCAGAGTTGTTGCGCAGCAGCAGCAGCAATTAA 1637
DB 2160 GGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2219
QY 1638 TCCCTCTAGCAATGAGATTTGCCAAGAGAGCTGAGGAAAGTCAAGCTTATGAGATT 1697
DB 2220 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA 2279

QY 1698 TCAAGAGAAAGATGAGAGAGCTTTGTGAGAGAGGAGAGATGATTAAGATCAAGA 1757
DB 2280 GCAGCAGAGAGCAGCAGCAGATGAGAGAGCAGCAGAGATGAGATCAAGA 2339
QY 1758 GAAAGAGATGAGAACTATGAAGAAGAGCATCAAGAGAGATTTGATCTGAGAGAAAGA 1817
DB 2340 GCAGAGTTAGAGCAGCAGCAGAGATTTAGAGAGCAGCAGAGATTTAGAGAGCA 2399
QY 1818 ATTTGATGAGCTTTGGAACAG 1839
DB 2400 GAGCAGAGATTTAGAGAGCAG 2421

RESULT 5
US-09-298-568-1
Sequence 1, Application US/09298568
Patent No. 6322792
GENERAL INFORMATION:
APPLICANT: Kieff, Elliott D.
APPLICANT: Ballester, Mary E.
APPLICANT: Kaye, Kenneth M.
TITLE OF INVENTION: RHADINO VIRUS IANA ACTS IN TRANS ON A UNIT OF RHADINO
FILE REFERENCE: 16412-10001R
CURRENT APPLICATION NUMBER: US/09/298,568
EARLIER FILING DATE: 1999-04-21
EARLIER APPLICATION NUMBER: US 60/109,422
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 1
LENGTH: 3489
TYPE: DNA
ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-1

Query Match 3.1%; Score 58.8; DB 4; Length 3489;
Best Local Similarity 43.4%; Pred. No. 6.7e-07;
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

QY 1218 GAAGCAAGATCTGGACATATTCATCAACTCTCAAGGCAAAACAAGGCTGAATTGCA 1277
DB 1800 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1859
QY 1278 GTTGAATCATACCAAGAGATGTTGTAAGAGAGCTGAGCAGATCTGAGAGCAATCA 1337
DB 1860 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1919
QY 1338 GCAGCTGAATCTTTAAGAACAGCTCTCAAAAACAAGCAGCAGCAAGGCTGTTGA 1397
DB 1920 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 1979
QY 1398 GGAATCTCTGGAATTTATGAGCAGAGAGCTGCTAGAACACTGCAAGAGATTAATCGATCGT 1457
DB 1980 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2039
QY 1458 GAGACAGAGAACTAAGATGACATGCAAGAACAGAGGAGAGATGAGCAGCAG 1517
DB 2040 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2099
QY 1518 GTTTTCATGATTCATCAATCAACAGATCCATGAAAGAGAGCAGCAAGAGAGATTT 1577
DB 2100 GAGACAGCAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 2159
QY 1578 CGAGATGTTGACAGCAGCAGAACTGTCAGAGTTGTTGCGCAGCAGCAGCAGCAATTAA 1637
DB 2160 GGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2219
QY 1638 TCCCTCTAGCAATGAGATTTGCCAAGAGAGCTGAGGAAAGTCAAGCTTATGAGATT 1697
DB 2220 GCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 2279
QY 1698 TCAAGAGAAAGATGAGAGAGTTTGTGAGAGAGAGATGCTGATTAAGATCAAGA 1757

QY	1458	GAGCAGAGAACTTAAGATGCAAGCTAAGAAACAGAACAGATGATGCAACAGACAG	1517
Db	19957	GCACAGGATGAGCAGACAGACGAGTGGCAGCAGCAGAGATGACAGACAGCAGATGACA	19898
QY	1518	GTTTTTCATGATTCATCAATCAACAGATCCATGAAAGAGAGACGCAAGAGAGAGATTT	1577
Db	19897	GGAGCAGACAGGATATGACAGGAGCAGCAGATGAGCAGACAGACACAGATGAGCAGACA	19838
QY	1578	CGAATGTTGACGACGACGAGAAACCTGCCAAGTTGTTGGCCAGCAGCAGCAACATTAA	1637
Db	19837	GGATGACACAGACGACGAGATGAGCAGCAGCAGAGATGACACAGCAGCAGCAGATGA	19778
QY	1638	TCCCTTAAGCAATTCAGATTTGCCGAAAGAGAGCGTGAGGATGTCGAAGTTCAATGCA	1697
Db	19777	GCACACAGCAGGATGACGACGACGACGAGATGAAACAGACAGCAGCAGAGAGCAGGA	19718
QY	1698	TCAAGAGAAAGAGATGAGAGAGTTTGTGGAAGAGAGAGATCTGTATAAAGATCAAGA	1757
Db	19717	GCACAGAGAGAGCAGGAGCAGAGATTAGAGAGCAGAGCAGACAGATTTAGAGATTAGGA	19658
QY	1758	GAGAGAGATGAGACATGAAGAAAGCATTCAAGAGAGATATTTGATCTGGAGAAAGA	1817
Db	19657	GCAAGAGTTGAGAGACGAGAGCAGGATTAGAGAGCAGAGCAAGATTAGAGAGCA	19589
QY	1818	ATTGATGAGGCTTTGGAACAG	1839
Db	19597	GGAGCAGAGATTAGAGAGACAG	19576

RESULT 8
US-08-757-669A-20/c
; Sequence 20, Application US/08757669A
; Patent No. 6183751

RESULT &

```

1  GENERAL INFORMATION:
2  APPLICANT: Chang, Yuan
3  APPLICANT: Bohenzky, Roy A.
4  APPLICANT: Russo, James J.
5  APPLICANT: Edelman, Isidore S.
6  APPLICANT: Moore, Patrick S.
7  TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS
8  TITLE OF INVENTION: SEQUENCES AND USES THEREOF
9  NUMBER OF SEQUENCES: 20
10 CORRESPONDENCE ADDRESS:
11 ADDRESSEE: Cooper & Dunham LLP
12 STREET: 1185 Avenue of the Americas
13 CITY: New York
14 STATE: New York
15 COUNTRY: U.S.A.
16 ZIP: 10036
17
18 COMPUTER READABLE FORM:
19 MEDIUM TYPE: floppy disk
20 COMPUTER: IBM PC compatible
21 OPERATING SYSTEM: PC-DOS/MS-DOS
22 SOFTWARE: PatentIn Release #1.0, Version #1.30
23 CURRENT APPLICATION DATA:
24 APPLICATION NUMBER: US/08/757,669A
25
26 FILING DATE:
27 CLASSIFICATION: 424
28 ATTORNEY/AGENT INFORMATION:
29 NAME: White, John P.
30 REGISTRATION NUMBER: 28,678
31 REFERENCE/DOCKET NUMBER: 45185-F
32 TELECOMMUNICATION INFORMATION:
33 TELEPHONE: (212) 278-0400
34 TELEFAX: (212) 391-0525
35
36 INFORMATION FOR SEQ ID NO: 20:
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 32207 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: double
41 TOPOLOGY: linear
42
43 MOLECULE TYPE: DNA (genomic)
44
45 US-08-757-669A-20

```

Query Match	3.1%;	Score 58.8;	DB 3;	length 32207;
Best Local Similarity	43.4%;	Pred. No. 2.5e-06;		
Matches 270;	Conservative	0;	Mismatches 352;	Indels 0;
			Gaps	0;

Qy	1218	GAAGCAGAGACTCTGGACATATTCATCAACAACCTCTCAAGCCAAACAAAGGCTCAAAATTCGA	1277
Db	20197	GCACGAGATGACACGACGAGGATGAGCAGCAGCAGATGACGACGACGACGAGATGAGCA	20138
Qy	1278	GTTGAATTCATVACCAAGAGTGTGTTAAAGAGCTTGAGCGAGATCTCTGAGACATCA	1337
Db	20137	GCAGCAGAGATGACACGACGAGGATGAGCAGCAGCAGATGAGCAGCAGCAGACGATGAGCA	20078
Qy	1338	GCAGCTGAACTTACTTTTAAGAACAAAGCTCTCAAAACGAACCAAGCCAGCAAGTGTCTTA	1397
Db	20077	GCACGACGAGATGACGACGAGGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGAGATGAGCA	20018
Qy	1398	GGAATCTCTGAAATTTATGAGCCAGAAAGCTGCGTAGMACTGCAAGAGATATTCGATCGT	1457
Db	20017	GCAGCAGAGATGACGACGAGGATGAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA	19958
Qy	1458	GAGACGAGAACTTAAGATGTCAGATGAACGAACGAGGAAGATGATGCAACAGCAG	1517
Db	19957	GCACGACGAGATGACACGACGAGGATGAGCAGCAGGATGAGCAGCAGCAGCAGAGATGAGCA	19898
Qy	1518	GTTTTTCATGAGATTCATCAATCAACAGATCCATGAAAGAAAGACGCAAGAGAGAGATTT	1577
Db	19897	GCAGCAGCAGAGATGAGCAGGAGCAGCAGATGAGCAGGACGACGACAGATGACGACAGCA	19838
Qy	1578	CGAGATTTTGCAGCAGCAGGAAAGTGCCAAAGTTGTTGGCCAGCAGCAGCAGAACTTAA	1637
Db	19837	GCATGAGCAGCAGCAGCAGGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCA	19778
Qy	1638	TCCCTCTAGCATGATGAGATTGCCGAAGAAGACCTGAGAAAGTGTCAAGCTTATGAGATT	1697
Db	19777	GCAGCAGCAGCAGAGATGAGCAGCAGCAGCAGATGAACGAGGACGACGAGAGCAGAGCA	19718
Qy	1698	TCAAGAGAAAGAGATGAGAGGATTTTGTGAAAGAGACGAGAGATGCTGATTTAAAGATCAAGA	1757
Db	19717	GCAGCAGGAGGACGACGAGAGATTTAAGAGACGAGGACGAGAGATTTAGAGATCTAGGA	19658
Qy	1758	GAAAGAGATGAGAGACATGAGAAAGAGGACATCAAGAGGATTTTGAATCTGAGAAAGA	1817
Db	19657	GCAGGAGTTTGAAGACAGGAGCAGGATTTAGAGAGCAGAGACGAGAGATTTAGAGAGCA	19598
Qy	1818	ATTGATGAGGCTTTGGAACAG	1839
Db	19597	GGAGCAGGAGTTAGAGAGCAG	19576

RESULT 9

```

US-09-230-371A-20/C
: Sequence 20, Application US/09230371A
: Patent No. 6348586
:
: GENERAL INFORMATION:
: APPLICANT: Chang, Yuan
: APPLICANT: Bohenzky, Roy A
: APPLICANT: Russo, James J
: APPLICANT: Edelman, Isidore S
: APPLICANT: Moore, Patrick S
:
: TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
: FILE REFERENCE: US97000001
: FILE REFERENCE: 45185-G-PCT-US
: CURRENT APPLICATION NUMBER: US/09/230,371A
: CURRENT FILING DATE: 1999-11-17
: PRIOR APPLICATION NUMBER: PCT/US97/13346
: PRIOR FILING DATE: 1997-07-22
:
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 20
:
: LENGTH: 32207
:
: TYPE: DNA
:
: ORGANISM: Kaposi's sarcoma-associated herpesvirus

```

US-09-230-371A-20

Query March	3.1%	Score 58.8;	DB 4;	Length 32207;
Best Local Similarity	43.4%	Pred. No. 2.5e-06;		
Matches 270; Conservative	0;	Mismatches 352;	Indels 0;	Gaps 0;

[illegible]

RESULT 10
 US-08-781-891-208/c
 ; Sequence 208, Application US/08781891
 ; Patent No. 6090620
 ; GENERAL INFORMATION:
 APPLICANT: Fu, Yang-Hui
 APPLICANT: Yu, Chang-En
 APPLICANT: Oshima, Junko
 APPLICANT: Mulligan, John T.
 APPLICANT: Schellenberg, Gerald D.
 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 TITLE OF INVENTION: WERNER'S SYNDROME
 NUMBER OF SEQUENCES: 209
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk

```

COMPTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620tenburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 208:
SEQUENCE CHARACTERISTICS:
LENGTH: 16442 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

```

Query Match	2.8%	Score 53.2;	DB 3;	Length 16442;
Best Local Similarity	47.6%	Pred. No. 7.6e-05;		
Matches 157; Conservative	0;	Mismatches 173;	Indels 0;	Gaps 0

[illegible]

US-09-618-166-208/c
Sequence 208, Application US/09618166
Patent No. 6583112
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
Yu, Chang-Shn
Oshima, Junko
Mulligan, John T.
Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed Intellectual Property Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDION TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: 166
APPLICATION NUMBER: US/09/618,166
FILING DATE: 17-Jul-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052.419CL
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 208:
SEQUENCE CHARACTERISTICS:
LENGTH: 16442 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-618-166-208

Query Match 2.8%; Score 53.2; DB 4; Length 16442;
Best Local Similarity 47.6%; Pred. No. 7.6e-05;
Matches 157; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1472 AGATGACATGACGAAGAGAGAGATGATGACACGACAGTTTTCATGATG 1531
DB 16437 AG 16378
QY 1532 CATCAACAGATCCATGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1591
DB 16377 AGCAG 16318
QY 1592 AGCAG 1651
DB 16317 AG 16258
QY 1652 AGCAG 1711
DB 16257 AG 16198
QY 1712 TGAGAGAGTTTGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
DB 16197 AG 16138
QY 1772 ACATGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1801
DB 16137 AGAT 16108

RESULT 12
US-08-781-891-209/c
Sequence 209, Application US/08781891
Patent No. 6090620
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSER: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA: 891
APPLICATION NUMBER: US/08/781,891
FILING DATE: 27-DEC-1996
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: No. 6090620cemburg Ph.D., Carol
REGISTRATION NUMBER: 39,317
REFERENCE/DOCKET NUMBER: 240052.419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-781-891-209

Query Match 2.7%; Score 49.8; DB 3; Length 51259;
Best Local Similarity 50.5%; Pred. No. 0.0015;
Matches 146; Conservative 0; Mismatches 142; Indels 1; Gaps 1;

QY 1525 ATGAGATTCATCAACAGATCCATGAAAGAGAGAGAGAGAGAGAGAGAT 1583
DB 290 ATGCTTAAACCAACCAATATCATGAAAGAGAGAGAGAGAGAGAGAGAG 231
QY 1584 GTTGACAGCAG 1643
DB 230 GAG 171
QY 1644 TAGCAATGACGATTCGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1703
DB 170 GCAG 111
QY 1704 GAAAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1763
DB 110 GAG 51
QY 1764 GATGAAGACATGAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATAT 1812
DB 50 GAG 2

RESULT 13
US-09-618-166-209/c
Sequence 209, Application US/09618166
Patent No. 6583112
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
APPLICANT: Yu, Chang-En
APPLICANT: Oshima, Junko
APPLICANT: Mulligan, John T.
APPLICANT: Schellenberg, Gerald D.
TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSER: Seed Intellectual Property Law Group
STREET: 701 Fifth Avenue, Suite 6300
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/618,166
FILING DATE: 17-Jul-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: McMaisters, David D.
REGISTRATION NUMBER: 33,963
REFERENCE/DOCKET NUMBER: 240052,419C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 209:
SEQUENCE CHARACTERISTICS:
LENGTH: 51259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 209:
US-09-618-166-209

Query Match 2.7%; Score 49.8; DB 4; Length 51259;
Best Local Similarity 50.5%; Pred. No. 0.0015;
Matches 146; Conservative 0; Mismatches 142; Indels 1; Gaps 1;
Db 1525 ATGATTCATCAACACAGATCCATGAAAGAA- GACGCCAAAGAGAGAAATTCGAGAT 1583
QY 290 ATGTCTAAACCAACAAATACATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 231
QY 1584 GTTCGACGACGAGAAAGTGCACAGATTGTTGGCCACAGACAGACAGACACTTAAATCCCTC 1643
Db 230 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 171
QY 1644 TAGCAATGACGATTCGCCAAGAGAGCTGAGAGAAAGTGTCAAGCTTCAGCA 1703
Db 170 GCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 111
QY 1704 GAAAGAGATGGAAGAGTTGTGGAAGAGAGAGAGATGCTGATTAAGATCAAGAGAA 1763
Db 110 GAGGGGACAG 51
QY 1764 GATGGAAGCATGAGAGAGAGAGAGATACAGAGAGATTTTGTGCTCGAG 1812
Db 50 CGAGGAG 2

RESULT 14
US-09-177-325-2
Sequence 2, Application US/09177325B
Patent No. 6214983
GENERAL INFORMATION:
APPLICANT: Tureci, Ozlem
APPLICANT: Sakin, Ugur
APPLICANT: Pfreundschuh, Michael
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins, A
TITLE OF INVENTION: Theoret
FILE REFERENCE: LUD 5525
CURRENT APPLICATION NUMBER: US/09/177,325B
CURRENT FILING DATE: 1998-10-22
NUMBER OF SEQ ID NOS: 15
SEQ ID NO 2
LENGTH: 1276
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-177-325-2

Query Match 2.5%; Score 47.4; DB 3; Length 1276;
Best Local Similarity 47.2%; Pred. No. 0.00088;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
Db 1516 AGGTTTTCATGATTCATCAACAGATCCATGAAGAGAGAGAGAGAGAGAT 1575
QY 607 ATGCTTAAGATATCATCGCAGAGAGAGAGAGAGATGATGACAGAGAGAGAGAG 666

QY 1576 TTGAGATGTTGACGACGAGAAAGTCCAGAGTTGTCGACGACGAGAGAGACATT 1635
Db 667 CCGCAGACGACGAGCCGAGAGCCGACGACGACAGAGCCGACGACGAGAGAGAG 726
QY 1636 AATCCCTTACCATATGACGTTGCCAAGAGAGCTGAGAGAGTGTCAAGCTTCATCAG 1695
Db 727 TAATAGTAGTAAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
QY 1696 TTCAAGAGAAAGAGATGAGAGCTTTGTGAGAGAGAGAGAGAGAGAGAGATTAAGATCA 1755
Db 787 AGAAG 846
QY 1756 GAGAGAGATGAGAGATGATGAAGAGAGAGATGATGATGATGATGATGATGATGATG 1815
Db 847 GAG 906
QY 1816 GAATT 1820
Db 907 AATT 911

RESULT 15
US-09-411-812A-2
Sequence 2, Application US/09411812A
Patent No. 6261778
GENERAL INFORMATION:
APPLICANT: Tureci, Ozlem
APPLICANT: Sakin, Ugur
APPLICANT: Pfreundschuh, Michael
TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode SCP Proteins, And
TITLE OF INVENTION: Theoret
FILE REFERENCE: LUD 5525.1 CIP
CURRENT APPLICATION NUMBER: US/09/411,812A
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: US 09/177,325
PRIOR FILING DATE: 1998-10-22
NUMBER OF SEQ ID NOS: 14
SEQ ID NO 2
LENGTH: 1276
TYPE: DNA
ORGANISM: Homo sapiens
US-09-411-812A-2

Query Match 2.5%; Score 47.4; DB 3; Length 1276;
Best Local Similarity 47.2%; Pred. No. 0.00088;
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;
Db 1516 AGGTTTTCATGATTCATCAACAGATCCATGAAGAGAGAGAGAGAGAT 1575
QY 607 ATGCTTAAGATATCATCGCAGAGAGAGAGAGATGATGACAGAGAGAGAGAG 666
Db 1576 TTGAGATGTTGACGACGAGAAAGTCCAGAGTTGTCGACGACGAGAGACATT 1635
QY 667 CCGCAGACGACGAGCCGAGAGCCGACGACGACAGAGCCGACGACGAGAGAGAG 726
Db 1636 AATCCCTTACCATATGACGTTGCCAAGAGAGCTGAGAGAGTGTCAAGCTTCATCAG 1695
QY 727 TAATAGTAGTAAGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 786
Db 1696 TTCAAGAGAAAGAGATGAGAGCTTTGTGAGAGAGAGAGAGAGAGAGAGATTAAGATCA 1755
QY 787 AGAAG 846
Db 1756 GAGAGAGATGAGAGATGATGAAGAGAGAGATGATGATGATGATGATGATGATGATG 1815
QY 847 GAG 906
Db 1816 GAATT 1820
QY 907 AATT 911

Wed Mar 31 08:14:28 2004

us-10-030-829-2.rnt

Page 9

Search completed: March 30, 2004, 03:15:08
Job time : 154.3 secs

This Page Blank (uspto)

QY 582 CCATGGATCACCGAAAGC

Db 572 TTTTAAAGCTGAAAAAGAAACAGTGTGTTTAAAGATTTGAAATGAAATACAT 631
 Qy 642 GTGATCGAGCAGATTAATGAACCAAGAGCAGTGGATTTGTCAGTTCGAAACG 701
 Db 632 GAGTTTGAACAAATTAATGAACAACTCGCAATGGAATGCCCCGCAATGAAATG 691
 Qy 702 ACCGTGTCATCGATGTTGTTATTA---CCTGACCCCTACTACTAGCTCAATGAGACAA 758
 Db 692 ACCGTGTCATCGATGTTGTTATTA---CCTGACCCCTACTACTAGCTCAATGAGACAA 751
 Qy 759 AGAGCTAGAGCAGATTAATGAAATGCTGAGATTTGAAATGAAATGCTGAA 818
 Db 752 GGGTTCTAAGAGGTTAACTTCAAGGAAATTTGGCTGATGCTGAGAGAGAGCTTTC 811
 Qy 819 GATGAGAGCGCATCTGTCTATTCCTGTGTGAGATTTTATGAGGAGAGGTTTGG 878
 Db 812 TCGCAGAGGGAATCTTCAAGTGTACAGAGGAGTGAACATTTGGAATGAAAGAGCTG-- 869
 Qy 879 TGAGATGAAAAGATTTATGAAATTTGCTGAGCTCCATGCTCATCATGATGATGAG 938
 Db 870 -CAAGAAAGCATGATCTGAGATGATGAGTGGCCCAATGGTTATTTATGAAATCCTT 928
 Qy 939 ACTGATTAAGAGCAGATTAATGATGCTGAGCTGAGCAGCAGCAGCAGCAGCAGCAG 998
 Db 929 CTGGAAGAAAGATGAAGATGATTAAGTGAAGGAGCAGTGAACCAAGAGCTCTTGA 988
 Qy 999 CTTCGACATGATGAGGCTCTTGAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1058
 Db 989 TTTTGAAGATTAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1048
 Qy 1059 GATGAGTGTCTGATGTTTGAAGAGCAGTGGCTGCTATTTGAGAGCGAAGCTTCCA 1118
 Db 1049 TATGAGGCTGTATATTTGAAAGCTCAGCTGTGGCTGATGAGAGCTGAAAGCTCTGA 1108
 Qy 1119 CCGGAGTGTGCTGAGATGAGGCTGATGATGATGATGATGATGATGATGATGATGATG 1175
 Db 1109 TAAACCTTTGTTATCAAGGTACAGCAGAGATTCAGAGCAGCAGCAGCAGCAGCAG 1168
 Qy 1176 GTTTTGAAGGAGTGTGAGCAGCAGTGTGAGCTTCTTGAACGAAAGCAGATCTGAG 1235
 Db 1169 TGTGCTGTGTGAGGAAAGGCACTATATGTTTCTTGAACAAAGAGATGAGAGG 1228
 Qy 1236 ATTCAATCAACACTCTCAAGCAGAAACAGGCTGAAATTTGAGTGAATCATACAGAG 1295
 Db 1229 ATTTAAACAGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1288
 Qy 1296 GATGTTTAAAGAGCTGAGCAGCAGATCTGAGAGCAATCAGAGCTGAACTACTTAA 1355
 Db 1289 GATGTTGATGATTAATGAAGAGATGATGATGATGATGATGATGATGATGATGATG 1348
 Qy 1356 GAACAGCTCTCAAAACAGAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1415
 Db 1349 GAACAGATGTTTAAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1408
 Qy 1416 GAGCAGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1475
 Db 1409 TACCAGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1468
 Qy 1476 GAGAGTGAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1535
 Db 1469 GAGAGTGAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1528
 Qy 1536 CAACAGATCTCAAAACAGAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1595
 Db 1529 TGAGATATTTCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1588
 Qy 1596 GGAAGCTGCAAGGTTGTTGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1655
 Db 1589 AGAGCCCTCAAGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1639
 Qy 1656 TTGCCAAGAGAGCTGAGAGAGTGTCAAGCTTCATGAGTTTCAAGAGAGAGAGAGAG 1715

Db 1640 TCGTCAGCTTAAGAGAAATATGTCAGAGATTCATTTGATTCAGAGTCAAGATGTG 1699
 Qy 1716 GAGATTTTGAAGAGAGAGAGAGATGCTGATTAAGATCAAGAGAGAGAGAGAGAT 1775
 Db 1700 AGATTTGAAGTTGAGCAGAGATGAGCTGATCAAGTTCACAGAGCAGAGAGATG 1759
 Qy 1776 GAGAGAGAGCAGCAGCAGAGATATTTGATCTGAGAGAGAGAGATTTGATGAGCTT 1835
 Db 1760 CAGAGAGAGATCATGATGATGATGAGCTGAGAGCTGAGAGAGAGATTTGAGAG 1819
 Qy 1836 ACAGCTCATGTACAGCAT 1854
 Db 1820 GGGCTGATGAGAGAGCAT 1838

RESULT 2
 US-10-424-599-90522/c
 ; Sequence 90522, Application US/10424599
 ; Publication No. US20040031072A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa Thomas J
 ; APPLICANT: Kovalic David K
 ; APPLICANT: Zhou Yihua
 ; APPLICANT: Cao Yongwei
 ; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53223)B
 ; CURRENT APPLICATION NUMBER: US/10/424,599
 ; NUMBER OF SEQ ID NOS: 285684
 ; SEQ ID NO 90522
 ; LENGTH: 2218
 ; TYPE: DNA
 ; ORGANISM: Glycine max
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)..(2218)
 ; OTHER INFORMATION: unsure at all n locations
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT3847_52754C.1
 ; US-10-424-599-90522

Query Match 21.3%; Score 400.4; DB 12; Length 2218;
 Best Local Similarity 67.1%; Pred. No. 1.4e-105; Indels 10; Gaps 3;
 Matches 615; Conservative 0; Mismatches 291;

Qy 455 AGAGTTTCTGACGTCGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 514
 Db 2153 AAGAGTCCCTTATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 2094
 Qy 515 ATGCTTTGATGAT-TCGTATGACGACTTGAAGATGATGATGATGATGATGATGATGAT 573
 Db 2093 ATGCTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2034
 Qy 574 CAAAAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 633
 Db 2033 CAAAAGCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1974
 Qy 634 GATAGCTTGCATGAGCAGATTAATGAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 693
 Db 1973 GATAGCTTGCATGAGCAGATTAATGAACAGCAGCAGCAGCAGCAGCAGCAGCAG 1914
 Qy 694 CAGAACGACCTGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 750
 Db 1913 CAGAGTGTCCCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1854
 Qy 751 AGAGCAAAAGAGCTGAGCAGATTAAGCTCATGAGAAATGAGCTGAAAGTTTGAAGAG 810
 Db 1853 AAAACAAAGGCTCAAAAGGATGAGATCATAGGAGATTTGCTGAGCTTTTGAAGAG 1794
 Qy 811 GATCTACAGATGAGAGCAGCATGTCATCTTGTGTGAGATTTATGAGCAGCTGAGAG 870
 Db 1793 GAACTGCGAGAGAGGCTGCGAGTATTCACCTGAGAGATTTGATGATGAGAG 1734


```
QY 871 GGTTCGGTGGAGTGAAGAAAGATTGTAATTTGCTGGCTCCAAATGTCATCATCAG 930
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1733 GGTTTAAAGATGAGGAAAGATCATGAAATGTTTGGCTCCGANTGGTTGATTCAG 1674
QY 931 AATACGACTGATGAAGACATTAACGATAAGTGGCTGGCAATGGGCAACCAAGAGCTG 990
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1673 AATACAGGCTTGAAACAGATGAAATGATTAAGTGGCTGAGTATGGGTAACCAAGAGCTT 1614
QY 991 CTGGAATACCTTGACAAAGTATGAGCTCTTAGAGCAAGCCATTCATAGTGTCCAGAGGC 1050
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1613 CTTAATATTTTAAAGCCATATGATGCTGTGAAGCTGCAGACGCTTAATGAGCCCAAGGGT 1554
QY 1051 CATTCGGAGTGAATGTTCTGATGTTTGAAGACAGTGGCTGCTATTTGAGGCGGAA 1110
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1553 CATCGGGGTTGAGCGCTTTGATTAATTTGAAGCATCAAGCTATAGTTATCTTGAGGCTGAG 1494
QY 1111 CGGCTTCACCGGAGTTAGCTGAGATGGGGTTAGATGAATTTGCTGGGGTCAGAAAGCCG 1170
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1493 CGCTACACAGCATTTGGCAGAAACAGAACTGACCGAGATGCTTGGTTAGTCAACAC 1434
QY 1171 AGATGTTTTT-----CTGAGAGTGTTCGCCAATGTAAGGCTTCTTGCAAGAACAA 1224
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1433 CGTAGATTAATTTCTCCCTGGTGGCCGCGCAGACGCTTATGATTAACATGGCAATMAAAGA 1374
QY 1225 GATTCGACATATTCATCACTCTCAAGGCAAAACAAAGCTGAAATTCGAGTTGAAA 1284
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1373 GACCTGACATTTTCAACAGACATTCGCAAGTAACTTAGATCTCAAAATGACATGAGA 1314
QY 1285 TCATACCAAGAGATGGTTGTAAGAGAGCTGAGGCAAGATCTTGAGAGCAATCAGAGCTG 1344
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1313 TCATACCAAGAGATGGTTGTAACCAACTTCGCAAGAAATGAGGCCAACAGAGCTT 1254
QY 1345 AACTACTTTAAGACA 1360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1253 AATTATTTGAAGAAA 1238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 3

```
US-10-424-599-91177
/ Sequence 91177, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 91177
/ LENGTH: 1377
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1
US-10-424-599-91177
```

```
Query Match 13.6%; Score 254.6; DB 12; Length 1377;
Best Local Similarity 66.3%; Pred. No. 3.6e-63;
Matches 417; Conservative 0; Mismatches 199; Indels 13; Gaps 3;
QY 352 TATGATTAACAATTGTGGACCCCACTGTATCTCGCCCTCTTGGAGAGAGATG 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29 TATGAAGCAATATGTGATCTCAAAATCTGTAAATTCGGCTCCTCACTGAGATGGCTG 88
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 412 AATGGCAGGCAAGAGAGATTTGCTCAGCAGACAGCTGTGAGAGATTTCTGACGTG 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 AACTGCATATCCAGACCTGGTGCATTCAGTCCAAATGTAAGGATGATAAATTTCCACGAG 148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
QY 472 G-----AGATGATGTGATAATGCTTCTGAGAAAGAAATGATTCGATGCTTGG 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 149 GACCTTCAAAAGAAATTAATGTGTGATGATGATGGGAGAGAGAAATCTGATGATTTG 208
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 523 GATGATTTCTGATGACACCTTGGCAAGTATGATTAATGACTCGGATGTGATCAAAGAGC 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 209 GAAGATATCTGATGATATCTGATGATGATGACGATGATTTCTGATCTAGTCAAAGAGT 268
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 583 CATGATCAAGAAAGAGATAATAGTGTCAAAGATTTCTTGGCAGCTTGGATGACTTGG 642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 269 CATGACACGTAAGAAAGCAAAATGTATTAAGAAATTTCTTGAATTTTGAATGGCTTG 328
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 643 TCATGACAGATTAATTAATCAACAGACAGAGCAATGCTTGCACCTTGTCAAGACGA 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 329 ACTGTTGAACAGATTAACCAACAGAAACAGTGGCATTTGTCAGCTTGCACAGAGTGGT 388
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 703 CTTGTGCTCATGATTTGATTAAC--CTGCACCTCTACTAGCTATATGCGAGACAAA 759
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 389 CTTGAGCTTATGATTTGATTTGATACAGAGAGCTGAGCCCTTGTGACACATGCCAAAGCAAAG 448
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 760 GGAGCTAAGCGAGTTAAGCTCATAGAGAAATGGCTGAAGTTTGAAGAAAGATCTACAG 819
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 449 GATCAAAAAGGTTGAAGATTCACAGGAGCTTGTATCTTTGATGAGATGAAATTAACG 508
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 820 ATGAGAGCGCATCTGTCAATCTTGTGTGATTAATGAGCAGTGAAGGTTTGGGT 879
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 509 AAAGGGTAACTTCAGTAATTCAGCTGGGGAAGTATTTGAAGTGAAGGTTTGAA 568
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 880 GAGGATGAAGAGATTTATGAATTTGCTGGCTCCATAGTCAATCAATGAATACTAGA 939
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 569 GAAGAG--AAAGACCATGATGATGATTTGGCTCCCAATGTTGTCAATCAGATACAAAG 627
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 940 CTGATTAAGCAGATTAAGATTAAGTGGCT 968
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 628 CTGAACAGATGAATATATAGCCGAT 656
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

RESULT 4

```
US-10-424-599-91176
/ Sequence 91176, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa Thomas J
/ APPLICANT: Kovalic David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
/ TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 91176
/ LENGTH: 1589
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
US-10-424-599-91176
```

```
Query Match 11.5%; Score 215.6; DB 12; Length 1589;
Best Local Similarity 64.6%; Pred. No. 9.2e-52;
Matches 358; Conservative 0; Mismatches 184; Indels 12; Gaps 2;
QY 352 TATGATTAACAATTGTGGACCCCACTGTATCTCGCCCTCTTGGAGAGAGATG 411
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29 TATGAAGCAATATGTGATCTCAAAATCTGTAAATTCGGCTCCTCACTGAGATGGCTG 88
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 412 AATGGCAGGCAAGAGAGATTTGCTCAGCAGACAGCTGTGAGAGATTTCTGACGTG 471
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 AACTGCATATCCAGACCTGGTGCATTCAGTCCAAATGTAAGGATGATAAATTTCCACGAG 148
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 472 G-----AGATGATGTGATAATGCTTCTGAGAAAGAAATGATTCGATGCTTGG 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

	Query Match	Similarity	Score	197.8	DB 12	Length	2890
	Best Local	Similarity	69.4%	Pred. No.	2,1e+66		
	Matches	284	Conservative	0	Mismatches	122	Indels 3; Gaps 1;
Qy	548	GTGATGATTATGACTCGATGTGATGTCAAAGACCCATGATGCAGAAAGCAGATTATGCT	607				
Db	2890	GTGATGATTATGATCTCGATACATACGTCAAATTAGCCACAGACTGTTAAGAAAGACCAAG	2831				
Qy	608	GGTTCAAAAAGTCTTTGGCAGCTTGGATAGCTTGTGCATGCAGCAGATPAATGAACCA	667				
Db	2830	GGTTTAAGATTTCTTTGAGAACCTTGGATGGCTTACTATTCGAAAGATCATGAAACGG	2771				
Qy	668	AGAGGCGTGGCATTTGCCAGCTTGTTCAGAAAGCAGCTGTCGATCGCATTTGGTATA---	724				
Db	2770	AAAAGCAGTGGCACTGTACAGCTTGTTCAGAGTGTCCCGGTGTTATTTGTTGTCACAG	2711				
Qy	725	ACCTGCACCTCTACTAGCTCATGCGAGCAAAAAGAGCTAGCGAGTTAAGTTCGCATA	784				

Query Match	9.3%;	Score 175.4;	DB 12;	Length 1103;
Best Local Similarity	65.7%;	Pred. No. 4e-40;		
Matches 285;	Conservative	0;	Mismatches 147;	Indels 2;
				Gaps 2;
QY	1172	GRATGTTTCTGAGGCTGTGCGCAACTGTAATGCGTTCCTTGCAACGAGCAAGATCTGG	1231	
Db	434	GTGGGCTCCCTGGGGGGAATGCCGAGCTCTATGATATACATGGCTCTCAAAAAGATATGG	375	
QY	1232	ACATATTCAATCAACACTCTCAAGGCCAAAACAAGCTGTAATTCGATGTGAATTCATC	1291	
Db	374	ACTTTTCAACAGACATTCGCCAAGTAAATCTGACTCAAAATATACATAAGATATATC	315	
QY	1292	AAGAG-ATGGTGTAAAGAGCTGAGGACGATCTCTGAGGACAAATCAGACGTGAATAC	1350	
Db	314	AGGAGAAATGGTTTATACCAAAATTCGGCAAAATGAATGAAGAACCAACAGCTGCTAT	255	
QY	1351	TTTAAACAAGCTCTCAAAAACAGAACCAAGCCGACAGTCTTGAGGAATCTCTGAA	1410	
Db	254	TTTAAACAAGGCTGTGAAAGGTGCAAAAACACAAAAGCTTGAAGAAATCTAATGGT	195	
QY	1411	ATTATGACGCAAGAGCTGGGTGAATCGACAGGATTAATCGATCGTGAGACAG-AGAAC	1469	
Db	194	ATAATGCTGAGAAGCTGCGAAGACAAATGAGGAAATTCGATTTGTAGGGAGGAGAAC	135	
QY	1470	TAAATGACGATGAACAGAACAGGAGAGATGATGACACGACAGGTTTTTCATGGA	1529	
Db	134	TAAATGCAACATGGAAGAGCCAAAAGNAGATGATATATGCAAGAACAAATTCCTTCAAGA	75	
QY	1530	TTCAATCAACAGATCCATGAAGAAGAACCGAAAGGAGAGAAATTTGAGATGTGCA	1589	
Db	74	CGAGATCAAAATCATTCATGATTCAAAGGATGCAAAAGAGAGAAAGATTTTCAGAGATGC	15	

QY 1590 GCAGCAGAACGTG 1603
DB 14 ACAGGAGGAACGTG 1

RESULT 7
US-10-424-599-73591/C
; Sequence 73591, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 73591
; LENGTH: 1375
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1375)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: PAT_MRT3847_37468C.1
US-10-424-599-73591

Query Match 9.0%; Score 169.8; DB 12; Length 1375;
Best Local Similarity 69.0%; Pred. No. 1.9e-38;
Matches 247; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

QY 548 GTGATGATTTGATCGATGTGATGCAAAAAGCCATGATCAGCAAGACAGATTAAGT 607
DB 1375 GTGATGATTTGATTTGATCTGATCTAGTCAGCAAGCCAGAGACTCGTAAGAGAGCAAT 1316
QY 608 GGTCAAAAAGTCTTTGGCAGCTTGATGATGCTTGTGATGAGCAGATTAATGAAC 667
DB 1315 GGTAAAGATTTCTTTGAGAACTTGATGATGCTTGTGATGAGCAATGAACCG 1256
QY 668 AGAGGAGTGGCATTGTCCAGCTTGTCAAGACGACCTGTCATCGATTTGTTA--- 724
DB 1255 AAAGGAGTGGCACCCTCAGCTTGTCAAGTGTCCGCTGCTATGATTTGTTAGACAG 1196
QY 725 ACCTGACCTCTACTAGCTCAGTGAAGGACAAAGAGCTAGCGAGTTAACTGCATA 784
DB 1195 GACTGACCTCTGATTTACTCATGCAAAACAAAGGGGTCAAAAGGGTGAAGATCATA 1136
QY 785 GAGATTTGCTGAAGTTTGAAGAGATCTACAGATGAAGGCGCATCTGTCATT 844
DB 1135 GGGAGTTTCTGAGCTTTTGAAGAGAACTGCGCAAAAGGGCTCTGAGTAATTCAC 1076
QY 845 GTGCTGAGATTTATGGCAGTGAAGGCTTTGGCTGAGATGAAGAGATTAATGAAT 902
DB 1075 CTGNGAAGATTTGTTGTAAGTGAAGGTTTAAAGATGAGGGAAGATCATGAAT 1018

RESULT 8
US-10-424-599-99770/C
; Sequence 99770, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 99770
; LENGTH: 951
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_61108C.1
US-10-424-599-99770

Query Match 5.8%; Score 109.6; DB 12; Length 951;
Best Local Similarity 54.5%; Pred. No. 5.7e-21;
Matches 313; Conservative 0; Mismatches 219; Indels 42; Gaps 3;

QY 1280 TGAATCATACCAAGATGTTGTAAAGAGCTGAGGAGATCTCTGAGCAATCAGC 1339
DB 949 TGAGATCATATCAAGATGTTGTGAAGAAAGATTAACATTTAATGATGATGTCGA 890
QY 1340 AGCTGAATCTTTAAGAACAGCTCTCAAAACAGAACAGCAGCAGGCTTGAG 1399
DB 889 AGCTGACTATTTATAGACATGTTAGCCAAAGAACAAATCAATATCAAGTGTGAC 830
QY 1400 AATCTCTGAAATTTATAGCAGAAAGCTCGTAGAATCTCAGAGATTAATCGATGTA 1459
DB 829 ACTCTTATGTAATATCAAGTGAAGTTAAGCATGACAACTGAAAAAACCGTGTGACC 770
QY 1460 GACAGAACTTAAGATGACATGAACAGAACAGGAGAGATGATGACACAGACGT 1519
DB 769 AACAGACA-----AGAGATGATGATGACCTAGAAAAA 737
QY 1520 TTTTCATGATTCATCAACAGATTCAGAAAGAGAGAGAGAGAGAGAGATTTG 1579
DB 736 ATTTCCAGGCCAAATCCAGACATTCACAGACCTACTGCAAGAGAGATTAAGTTTG 677
QY 1580 AGATGTTGATTAATC 1639
DB 676 TGAATTTACAGCTGAG-----GCAATGCAAGAAAGATGAAGAGTCTCTGAAAGAT 623
QY 1640 CCTTACAGATGACGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTC 1699
DB 622 CTTCTGAGAAAGAGC-----AAGTGAAGATATCTCTAGTTTCTGAAACCTTC 572
QY 1700 AAGAGAAAGATGATCAAGAGA 1759
DB 571 AAGACAGAGAGATGAAGCAATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATCATGA 512
QY 1760 AGAAGATGAGAGACATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTTGATCTGAGAGAGAT 1819
DB 511 AGAAAAAGTTGCTCTGAGT 452
QY 1820 TTGATGAGCTTTGAAACAGCTCATGTACAGCA 1853
DB 451 TAGAGATGAACTGACGAGCTCATGAGACAGTA 418

RESULT 9
US-09-294-093B-1296
; Sequence 1296, Application US/09294093B
; Patent No. US2001005135A1
; GENERAL INFORMATION:
; APPLICANT: Laligudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; PRIOR FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; NUMBER OF SEQ ID NOS: April 21, 1998
; SOFTWARE: PERL Program

```

; SEQ ID NO 1296
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700344080H1
; NAME/KEY: unsure
; LOCATION: 309
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-1296

```

Query Match 3.4%; Score 63.6; DB 9; Length 310;

Best Local Similarity 53.3%; Pred. No. 7.6e-08; Matches 163; Conservative 0; Mismatches 134; Indels 9; Gaps 1;

```

QY 1476 GCAGCATGAAACAGACGAGAGATGATGACGACGACGATTTTTCATGATTCAT 1535
DB 1 GAAGCATGAGATGACGAGAGAGATGAGATTCGAGAGAAATTTTCATGATTCAT 60
QY 1536 CAACACATTCATGAAAGAGAGACGCAAGGAAATTCGAGATTTGACAGACA 1595
DB 61 TGAGATATTCACAGAGCCACAGAGATTAAGAAAGCAGTTGAGAGATTTGACAGA 120
QY 1596 GGAACGTGCCAAGGTTGTGGCCAGCAGCAGCAGAACTTAATCCCTTAGCATGACGA 1655
DB 121 AGAGGCTCAAGGCCACAGACGCTTTGATG-----TTGATTTGGAATGAAAGA 171
QY 1656 TTCCGAAAGAGAGCTGAGAGATGTCAGCTTCATGATTCAGAGAAAGATGGA 1715
DB 172 TCGTAGCTAAGAGAGAAATGTGCGAAGTTGATTCATGATTCAGAGATGTCG 231
QY 1716 GGAATTTGTGAAAGAGAGAGATGTCATTAAGATCAAGAGAAAGATGGAAGCAT 1775
DB 232 AGAGTTTGAAGTTGAGCAGATGAGCTGATCAAGTCCACAGAGACAAAGATTGAAGCT 291
QY 1776 GAAGAA 1781
DB 292 CAAGAA 297

```

RESULT 10
US-09-894-273-1
; Sequence 1, Application US/09894273
; Publication No. US20040037847A1
; GENERAL INFORMATION:
; APPLICANT: Ballester, Mary E.
; APPLICANT: Kieff, Elliott D.
; TITLE OF INVENTION: KAYE, Kenneth M.
; TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
; FILE REFERENCE: 16412-10001R
; CURRENT APPLICATION NUMBER: US/09/894,273
; PRIOR FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/109,422
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-894-273-1

Query Match 3.1%; Score 58.8; DB 12; Length 3489;
Best Local Similarity 43.4%; Pred. No. 8.2e-06;
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

```

QY 1218 GAAGCAAGATCTGACATTTCAATCAACATCTCAAGCAAAACAAGCTGAATTGCA 1277
DB 1800 GCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1859
QY 1278 GTTGAATCATATCAAGAGATGTTGTAAGAGCTGAGGACGATCTCTGAGACATCA 1337

```

```

DB 1860 GCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1919
QY 1338 GCAGTGAATCACTTTTAAGAAACAAGCTTCAAAAACAACAGCAGCCCAAGTGTCTTA 1397
DB 1920 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1979
QY 1398 GGAATCTTGAATTTATGACCGAAGATGCGTGAATCTGACAGAGATTAATGATCGT 1457
DB 1980 GCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGATGAGCA 2039
QY 1458 GAGACAGAACTAAGATGACATGAAACAGAAACAGGAAAGATGATGACACAGCAG 1517
DB 2040 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 2099
QY 1518 GTTTTCATGATTCATCAACATCAACATGATCAAGAAAGACCAAGCAAGGAAATTT 1577
DB 2100 GAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 2159
QY 1578 CGAGATGTTGACAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1637
DB 2160 GATGAGCAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 2219
QY 1638 TCCCTTAGCATGACGATTTGCCAAGAGACTGAGAGAGTCAAGCTTCATGAGTT 1697
DB 2220 GCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 2279
QY 1698 TCAAGAAAGATGAGAGAGATTTGTGAAAGAGAGAGATGCTGATTAAGATCAAGA 1757
DB 2280 GCAGCAGAGAGAGAGAGAGATGAGCAGCAGATGAGCAGCAGATGAGCA 2339
QY 1758 GAAGAGATGAGAAACATGAAAGAGCCTCAAGAGAGATTTGATCTGAGAAAGA 1817
DB 2340 GCAAGATTTGAGAGAGAGAGAGAGATGAGCAGCAGATGAGCA 2399
QY 1818 ATTGATGAGCTTTGGAACG 1839
DB 2400 GAGCAGAGATTTAGAGAGCAG 2421

```

RESULT 11
US-10-294-804-1
; Sequence 1, Application US/10294804
; Publication No. US20030133948A1
; GENERAL INFORMATION:
; APPLICANT: Robertson, Kyle S.
; APPLICANT: Colter, Murray A.
; TITLE OF INVENTION: Methods to Inhibit or Enhance the Binding of Viral DNA
; TITLE OF INVENTION: to Genomic Host DNA
; FILE REFERENCE: UM-03778
; CURRENT APPLICATION NUMBER: US/10/294,804
; PRIOR FILING DATE: 2002-11-14
; PRIOR APPLICATION NUMBER: US/09/410,399
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3489
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-10-294-804-1

Query Match 3.1%; Score 58.8; DB 14; Length 3489;
Best Local Similarity 43.4%; Pred. No. 8.2e-06;
Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;

```

QY 1218 GAAGCAAGATCTGACATTTCAATCAACATCTCAAGCAAAACAAGCTGAATTGCA 1277
DB 1800 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1859
QY 1278 GTTGAATCATATCAAGAGATGTTGTAAGAGCTGAGGACGATCTCTGAGACATCA 1337
DB 1860 GCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 1919

```

QY 1338 GCAGCTGAATCTTTAAGAACAGCTCTCAAAAACAGAACACAGCCAGGCTGCTGA 1397
| | | | |
DB 1920 GCAGCAGATGAGCAGACAGAGATGACGACGAGATGACAGACAGATGAGCA 1979
| | | | |
QY 1398 GGAATCTTGGAATTTATGACGAGAACTGCTGATAGAACTGCAAGAGATTAATCGAT 1457
| | | | |
DB 1980 GCAGCAGATGAGCAGACAGATGAGCAGCAGATGAGCAGCAGATGAGCA 2039
| | | | |
QY 1458 GAGACAGAACTTAAGATGAGCATGAAACAGAGGAAAGATGATGACAGCAG 1517
| | | | |
DB 2040 GCAGCAGATGAGCAGACAGATGAGCAGCAGATGAGCAGCAGATGAGCA 2099
| | | | |
QY 1518 GTTTTCATGATTCATCAATCAACAGATCCATGAAAGAGAGCAGAAAGAGATTT 1577
| | | | |
DB 2100 GGAGCAGCAGATGAGCAGAGAGCAGAGATGAGCAGCAGCAGATGAGCAGCA 2159
| | | | |
QY 1578 CGAGATGTTGACAGCAGAGAACGTCGCAAGGTTGTCGACAGCAGCAGCAATTA 1637
| | | | |
DB 2160 GGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGA 2219
| | | | |
QY 1638 TCCCTTACCAATGAGATTCGCGAAGAGAGCTGAGAGATGCAAGCTTCAAGTT 1697
| | | | |
DB 2220 GCAGCAGCAGCAGATGAGCAGCAGCAGATGAAACAGAGCAGCAGAGAGCA 2279
| | | | |
QY 1698 TCAAGAGAAAGATGAGAGATTTGTGAAAGAGAGAGATGCTGATTAAGATCA 1757
| | | | |
DB 2280 GCAGCAGAGAGAGCAGAGCAGAGATTAAGAGCAGAGCAGAGATTAAGATCA 2339
| | | | |
QY 1758 GAGAAAGATGAGAGACATGAAAGAGGATCAAGAGAGATTTGATCTGAGAA 1817
| | | | |
DB 2340 GCAGAGTTAGAGAGCAGAGAGAGATTAAGAGAGCAGAGCAGAGATTAAGAG 2399
| | | | |
QY 1818 ATTGATGAGGCTTTGGAACAG 1839
| | | | |
DB 2400 GGAGCAGAGTTAGAGAGAGCAG 2421
| | | | |

RESULT 12

US-10-085-117-61/c
; Sequence 61, Application US/10085117
; Publication No. US2003023234A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Engelhard, Eric K.
; TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
; FILE REFERENCE: 529452000121
; CURRENT APPLICATION NUMBER: US/10/085,117
; CURRENT FILING DATE: 2002-02-27
; PRIOR APPLICATION NUMBER: US 09/798,586
; PRIOR FILING DATE: 2001-03-02
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 96602
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(96602)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-61

Query Match 3.1%; Score 58; DB 15; Length 96602;

Best Local Similarity 44.4%; Pred. No. 0.0001;

Matches 189; Conservative 0; Mismatches 236; Indels 1; Gaps 1;

QY 1447 AATCGATGCTGAGACAGACATTAAGATGACATGAGAACAGAGAGATGAT 1506
| | | | |
DB 51071 AAG 51012
| | | | |
QY 1507 GCACACGACAGGTTTTCATGATTCATCAAAACAGATCCATGAAAGAGAGCA 1565
| | | | |

DB 51011 GAGAGAGGGGGAANN 50952
| | | | |
QY 1566 GGAGGAGATTTTCAGATGTTGACAGCAGAGAACGTCGCAAGGTTGTCGACAGCA 1625
| | | | |
DB 50951 GGAG 50892
| | | | |
QY 1626 GCAGAACATTAATCCCTTACGATGACATGATGATGATGATGATGATGATGAT 1685
| | | | |
DB 50891 GGAG 50832
| | | | |
QY 1686 CTTTCATGCTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1745
| | | | |
DB 50831 GGAG 50772
| | | | |
QY 1746 AAAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTTGA 1805
| | | | |
DB 50771 GGAG 50712
| | | | |
QY 1806 TCTGAGAAAGATTTGATGAGGCTTTGAGAACGCTCATGTCAAGCATGCTTCACAA 1865
| | | | |
DB 50711 GAAAG 50652
| | | | |
QY 1866 TGAAGA 1871
| | | | |
DB 50651 AGAAGA 50646
| | | | |

RESULT 13

US-10-424-599-29202/c
; Sequence 29202, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 29202
; LENGTH: 473
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_12636C.1
US-10-424-599-29202

Query Match 2.9%; Score 53.8; DB 12; Length 473;

Best Local Similarity 53.6%; Pred. No. 7.1e-05;

Matches 112; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 1650 TGACGATTCGCGAAGAGAGCTGAGAGATGCAAGCTTCATGAGTTTCAAGAGAA 1709
| | | | |
DB 452 TGAGGATTAATGACATCAAGGTGAGAGAGATGATGAAATTTGTGAGCTCAAGATTCG 393
| | | | |
QY 1710 GATGAGAGAGTTTGTGAGAGAGAGAGATGCTGATTAAGATCAAGAGAAAGATGA 1769
| | | | |
DB 392 CTCGAGAACTTTGTTGCTGAAAGAGAGAACTGTTGATGCTCATGCGATGATTC 333
| | | | |
QY 1770 AGACATGAAAGAGAGAGATCAAGAGAGATTAATTTGATCTGAGAAAGATTTGATGAGC 1829
| | | | |
DB 332 TGCATTAAGAGCGAGCATTTGGAGAGAAAGATTAAGATGAGAGAGAGATTTGATGAGCA 273
| | | | |
QY 1830 TTTGGAACGCTCATGTCAAGCATGCC 1858
| | | | |
DB 272 ACTGCCAATCTCATGAGAAAGTACTCC 244
| | | | |

RESULT 14

US-10-374-077-208/c

Wed Mar 31 08:14:28 2004

us-10-030-829-2.rmpb

Page 8

```
Sequence 208, Application US/10374077
Publication No. US20040006779A1
GENERAL INFORMATION:
APPLICANT: Fu, Yung-Hui
            Yu, Chang-En
            Oshima, Junko
            Mulligan, John T.
            Schellenberg, Gerald D.
TITLE OF INVENTION: ANTIBODIES AGAINST GENE PRODUCTS RELATED TO
                    WERNER'S SYNDROME
NUMBER OF SEQUENCES: 209
CORRESPONDENCE ADDRESS:
ADDRESSER: Seed Intellectual Property Law Group
STREET: 701 Fifth Avenue, Suite 6300
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/374,077
FILING DATE: 25-Feb-2003
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Roseman, Stephen
REGISTRATION NUMBER: 43,058
REFERENCE/DOCKET NUMBER: 100107.401D1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 208:
SEQUENCE CHARACTERISTICS:
    LENGTH: 16442 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-10-374-077-208
Query Match      2.8%; Score 53.2; DB 15; Length 16442;
Best Local Similarity 47.6%; Pred. No. 0.0009;
Matches 157; Conservative 0; Mismatches 173; Indels 0; Gaps 0;
QY      1472 AGATGCAACATGAAACAGAACAGGAAAGATGATGACACACAGATTTCATGAGATT 1531
DB      16437 AGGAGGAGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16378
QY      1532 CAATCAACAGATCCATGAAAGAAAGACGCAAAAGAGAGAGAAATTTTCAGACTTC 1591
DB      16377 AGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16318
QY      1592 ACCAGGAACTGCTCAAGGTTTGGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16258
DB      16317 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16198
QY      1652 AGCATTTGCCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1711
DB      16257 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16198
QY      1712 TGGAGAGAGTTTGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1771
DB      16197 AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 16138
QY      1772 ACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1801
DB      16137 AGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 16108
```

RESULT 15

```
US-10-108-260A-1966
Sequence 1966, Application US/10108260A
Publication No. US20040005560A1
GENERAL INFORMATION:
APPLICANT: HELIX RESEARCH INSTITUTE
TITLE OF INVENTION: HI-A0106
FILE REFERENCE: HI-A0106
CURRENT APPLICATION NUMBER: US/10/108,260A
CURRENT FILING DATE: 2002-03-27
NUMBER OF SEQ ID NOS: 5458
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1966
LENGTH: 2479
TYPE: DNA
ORGANISM: Homo sapiens
US-10-108-260A-1966
Query Match      2.8%; Score 52.8; DB 15; Length 2479;
Best Local Similarity 43.9%; Pred. No. 0.00038; Indels 0; Gaps 0;
Matches 225; Conservative 0; Mismatches 287;
QY      1328 AGACATATCAGCAGCTGAACTCTTAAGAACAAAGCTCTCAAAACAGAACAGACGCCA 1387
DB      1377 AGGAGAAAGAGATGTGAGAGAGAGAGAGAGAGATTCAGACAGAGAGAGATACGAGAG 1436
QY      1388 AGGTGCTTGAAGAAATCTTGAAATTTAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1447
DB      1437 AGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1496
QY      1448 ATCCGATCTGAGACAGAGAACTTAAGATCCAGATGAAACAGAACAGGAGAGAGAGAG 1507
DB      1497 AGGAGATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1556
QY      1508 CACACAGAGGTTTTCATGATTCATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1567
DB      1557 AGAAGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1616
QY      1568 AGGAGAAATTTGAGAGATGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1627
DB      1617 AGGAGAGAGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1676
QY      1628 AGAATTAATTCCTCTTACATGACATTCCTGAAAGAGAGAGAGAGAGAGAGAGAGAG 1687
DB      1677 AGGAGAGAGAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1736
QY      1688 TCATGAGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1747
DB      1737 AGCAGAGAGAGAGATACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1796
QY      1748 AAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1807
DB      1797 GGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1856
QY      1808 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1839
DB      1857 TGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1888
```

Search completed: March 30, 2004, 08:58:17
Job time : 679.885 secs

CC 5GS3 results in plants with increased resistance to viruses, while
CC inactivation of 5GS3 in transgenic plants (e.g. by expressing antisense
CC RNA, by mutation or by homologous recombination) increases the level of
CC the transgene product. This product may e.g. impart resistance (to
CC herbicide, insects or pathogens), alter contents of essential fatty acids
CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or
CC interferon

XX Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 100.0%; Score 1878; DB 4; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGAGTTCTAGGGCTGTCCTCAATGCTCTAAGAAAAGAACCTTCAGGCTGCTTATAGGCT 60
DB 1 ATGAGTTCTAGGGCTGTCCTCAATGCTCTAAGAAAAGAACCTTCAGGCTGCTTATAGGCT 60
QY 61 GAGGTGAACAGTTGGTTCAAGTTTGGAGGGAGAGAGCTGGCTTCTTCAAGATGAT 120
DB 61 GAGGTGAACAGTTGGTTCAAGTTTGGAGGGAGAGAGCTGGCTTCTTCAAGATGAT 120
QY 61 GAGGTGAACAGTTGGTTCAAGTTTGGAGGGAGAGAGCTGGCTTCTTCAAGATGAT 120
DB 61 GAGGTGAACAGTTGGTTCAAGTTTGGAGGGAGAGAGCTGGCTTCTTCAAGATGAT 120
QY 121 GAGGAGAGTGGAGGCTATTTCCAAAGAAACAAAGACAAACAGGAAACCTTCTGGA 180
DB 121 GAGGAGAGTGGAGGCTATTTCCAAAGAAACAAAGACAAACAGGAAACCTTCTGGA 180
QY 181 AAAAATTGGGTTCTCAAGATTCCTCTAGAGCTTGGGGTGGTCAAGAGCAAGG 240
DB 181 AAAAATTGGGTTCTCAAGATTCCTCTAGAGCTTGGGGTGGTCAAGAGCAAGG 240
QY 181 AAAAATTGGGTTCTCAAGATTCCTCTAGAGCTTGGGGTGGTCAAGAGCAAGG 240
DB 181 AAAAATTGGGTTCTCAAGATTCCTCTAGAGCTTGGGGTGGTCAAGAGCAAGG 240
QY 241 AGAGGTGACACGATCTGAGAGAGAAACATGATCCGGAGAGGTAAACGCAATGAT 300
DB 241 AGAGGTGACACGATCTGAGAGAGAAACATGATCCGGAGAGGTAAACGCAATGAT 300
QY 301 CGGGGCAATTCAGCTAATCTGCTGCGGAGAGAGGTTGAGCAGAAAGATGATTAAC 360
DB 301 CGGGGCAATTCAGCTAATCTGCTGCGGAGAGAGGTTGAGCAGAAAGATGATTAAC 360
QY 361 AACTTTGGGCAACCCCACTGATCTGCTGCGGAGAGAGGTTGAGCAGAAAGATGAT 420
DB 361 AACTTTGGGCAACCCCACTGATCTGCTGCGGAGAGAGGTTGAGCAGAAAGATGAT 420
QY 421 GCAAGAGAGAGTTGCTCTCAGACACAGCTGTCAGAGAGTTCCGACGTCGAGATGAT 480
DB 421 GCAAGAGAGAGTTGCTCTCAGACACAGCTGTCAGAGAGTTCCGACGTCGAGATGAT 480
QY 481 GTGGAATATGCTTGAAGAGAGAGATGATCCGATGCTTGGATGATTCGATGACAC 540
DB 481 GTGGAATATGCTTGAAGAGAGAGATGATCCGATGCTTGGATGATTCGATGACAC 540
QY 541 CTTCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 541 CTTCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 541 CTTCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
DB 541 CTTCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
QY 601 AATTAAGGTTCAAAAAGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGAT 660
DB 601 AATTAAGGTTCAAAAAGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGAT 660
QY 661 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
DB 661 GAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
QY 721 TATAAAGGTTCAAAAAGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGATGAT 780
DB 721 TATAAAGGTTCAAAAAGTTCTTGGCAGCTTGGATGATGATGATGATGATGATGAT 780
QY 781 CATAGAGATGAGCTGAAGTTTGAAGAGAGATGATGATGATGATGATGATGATGAT 840
DB 781 CATAGAGATGAGCTGAAGTTTGAAGAGAGATGATGATGATGATGATGATGATGAT 840
QY 841 CTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
DB 841 CTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
```

```
QY 901 ATTGTGAGCTCCAAATGCTATCATATCATAGATTAAGTGAATGATGATGATGAT 960
DB 901 ATTGTGAGCTCCAAATGCTATCATATCATAGATTAAGTGAATGATGATGATGAT 960
QY 961 AAGTGGCTGGGCAATGGGCAACCAAGAGCTGAGATTAATCTTCAAGATGATGATGAT 1020
DB 961 AAGTGGCTGGGCAATGGGCAACCAAGAGCTGAGATTAATCTTCAAGATGATGATGAT 1020
QY 1021 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
DB 1021 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
QY 1081 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1081 AAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
QY 1141 TTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 TTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
QY 1201 TTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 TTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 ACAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 ACAAGGCTGAATTCGATGATGATGATGATGATGATGATGATGATGATGAT 1320
QY 1321 ATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
DB 1321 ATCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1380
QY 1381 CAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
DB 1381 CAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
QY 1441 GAGGATTAATCGGATTCGATGATGATGATGATGATGATGATGATGATGATGAT 1500
DB 1441 GAGGATTAATCGGATTCGATGATGATGATGATGATGATGATGATGATGAT 1500
QY 1501 ATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
DB 1501 ATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
DB 1561 GCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
QY 1621 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1621 CAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
QY 1681 TCAAGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
DB 1681 TCAAGCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
QY 1741 CTGATTAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
DB 1741 CTGATTAAGATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 TTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
DB 1801 TTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
QY 1861 CACAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
DB 1861 CACAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
```

RESULT 2
AAE25373 standard; DNA; 3275 BP.
ID AAF25373

QY 1068 TCTGATGTTGAGACAGTGGCCACTGGCTATTTGGAGGCGGAGCCCTCCAGCGGAGTT 1127
 DB 1836 TCTGATGTTGAGACAGTGGCCACTGGCTATTTGGAGGCGGAGCCCTCCAGCGGAGTT 1895
 QY 1128 AGCTGAGATGGGGTTAGATAGATTCCTGGGGTCAGAGCGCAGTATGTTTCTGGAGG 1187
 DB 1896 AGCTGAGATGGGGTTAGATAGATTCCTGGGGTCAGAGCGCAGTATGTTTCTGGAGG 1955
 QY 1188 TGTTCGCCAAGCTGATGAGCTTCCTTGGCAAGCAAGCAAGTCTGGCAATTCATCAACA 1247
 DB 1956 TGTTCGCCAAGCTGATGAGCTTCCTTGGCAAGCAAGCAAGTCTGGCAATTCATCAACA 2015
 QY 1248 CTCTCA----- 1253
 DB 2016 CTCTCAAGGTTCTCTCCCAAGAAATTTGATATGCTTTAGTTTGTGATTTGAAT 2075
 QY 1254 -----AGG 1256
 DB 2076 TTTAAGTTTGTGTCGGTTATGATCTGTTATGATATATCTATGATTCATTAGG 2135
 QY 1257 CAAGCAAGGCTGAATTTGAGTTGAATTCATCCAGAGATGTTGTAAGAGAGCTGAG 1316
 DB 2136 CAAGCAAGGCTGAATTTGAGTTGAATTCATCCAGAGATGTTGTAAGAGAGCTGAG 2195
 QY 1317 GCGATCTCTGAGGACATGACAGCTGAATCTTTAAGAACAGCTCTCAAAACAGAA 1376
 DB 2196 GCGATCTCTGAGGACATGACAGCTGAATCTTTAAGAACAGCTCTCAAAACAGAA 2255
 QY 1377 CAAGCAAGGCTGTTGAGGATCTCTGAAATTTATAGGAGAAAGCTGTGTAAGAC 1436
 DB 2256 CAAGCAAGGCTGTTGAGGATCTCTGAAATTTATAGGAGAAAGCTGTGTAAGAC 2315
 QY 1437 TGCAGAGATPATGCGATGCGAGACAGAGAACTTAAGATCAGACATGACAGAAACAGGA 1496
 DB 2316 TGCAGAGATPATGCGATGCGAGACAGAGAACTTAAGATCAGACATGACAGAAACAGGA 2375
 QY 1497 AG----- 1498
 DB 2376 AGAGTATGATTTTCTGAAATTCACAACTTGATTTGATTTACTACTGATTC 2435
 QY 1499 -----AGATGATGACACGA 1514
 DB 2436 CATTTTGTATTATTTGTCACAAACAAACCTGTGGTGGTTGAAGATGATGACACGA 2495
 QY 1515 CAGTTTTTTCATGATTCATCAACAGATCCATGAAAGAGAGACCGCAAGAGAGGAA 1574
 DB 2496 CAGTTTTTTCATGATTCATCAACAGATCCATGAAAGAGAGACCGCAAGAGAGGAA 2555
 QY 1575 TTTTCAGATGTTGCGACGACGAGAAAGCTGCCAAGTGTGTCAGACGACGACGAGAAAT 1634
 DB 2556 TTTTCAGATGTTGCGACGACGAGAAAGCTGCCAAGTGTGTCAGACGACGACGAGAAAT 2615
 QY 1635 TTAATCCTCTACAGATGAGATGCGGAAG----- 1665
 DB 2616 TTAATCCTCTACAGATGAGATGCGGAAG----- 1665
 QY 1666 ----- 1665
 DB 2676 CTGGCGTTTGTGTTTCAACCTAAGATGATGATTTCCGGTTTGAATCTTTG 2735
 QY 1666 -AGAGCTGAGAAAGTCAAGCTTCATCCAGTTTCAAGAGAAAGATGAGAGATTTGT 1724
 DB 2736 CAGAGCTGAGAAAGTCAAGCTTCATCCAGTTTCAAGAGAAAGATGAGAGATTTGT 2795
 QY 1725 GGAAGAGAGAGATGCTGATTAAGATCAAGAGAGAGATGAGAGACATGAAGAAG 1784
 DB 2796 GGAAGAGAGAGATGCTGATTAAGATCAAGAGAGAGATGAGAGACATGAAGAAG 2855
 QY 1785 GCATCAAGAGAGATTTGATTTGAGAAAGATTTGATGAGCTTTGAGACAGTCT 1844
 DB 2856 GCATCAAGAGAGATTTGATTTGAGAAAGATTTGATGAGCTTTGAGACAGTCT 2915

QY 1845 GTTACAGCATGGCTTCCACATGATGATGTA 1878
 DB 2916 GTTACAGCATGGCTTCCACATGATGATGTA 2949
 RESULT 3
 ABX81661
 ID ABX81661 standard; cDNA; 272 BP.
 XX
 AC ABX81661;
 XX
 DT 24-Apr-2003 (first entry)
 XX
 DE Corn ear-derived polynucleotide (cpd) #121.
 XX
 XX Corn ear-derived polynucleotide; cpd; cDNA library; SATMON022; SATMON023;
 KW structural gene; functional gene; regulatory gene; gene expression;
 KW corn ear-specific profile; gene transcription; plant breeding program;
 KW hybrid plant; desirable trait expression; growth; development;
 KW inheritance; desired characteristic; quality; yield;
 KW disease resistance; environmental adaptability; quality; yield;
 KW multigene trait; plant; gene; se.
 KW
 OS Zea mays.
 XX
 PN US6476212-B1.
 XX
 PD 05-NOV-2002.
 XX
 PF 14-MAY-1999; 99US-00313294.
 XX
 PR 26-MAY-1998; 98US-0086722P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 PA
 XX Laiqudi RV, Ito LY, Sherman BK;
 PI
 XX WPI; 2003-208840/20.
 DR
 XX
 PT Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with growth,
 PT development.
 PT
 XX
 PS Example; SEQ ID NO 121; 390bp; English.
 XX
 CC The present invention relates to the isolation of corn ear-derived
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022
 CC and SATMON023. Some of the cdps uniquely identify structural, functional,
 CC and regulatory genes of corn ear. The polynucleotide sequences are
 CC useful for detecting cdps in a sample, for producing a corn ear-specific
 CC profile of gene transcription, for detecting altered gene expression in
 CC hybrid or hybrid plants, and for screening several molecules for specific
 CC binding to the polynucleotide. The cdps are useful to identify, isolate,
 CC or extend identical or related corn-ear nucleic acid sequences from DNA
 CC libraries, and in nucleic acid amplification or hybridization techniques
 CC to follow the expression of desirable traits through plant breeding
 CC programs. Preferably, the cdps are used to identify, evaluate, alter, or
 CC follow the inheritance of desired characteristics associated with growth,
 CC and development, disease resistance, environmental adaptability, quality,
 CC and yield of corn. The cdps are also useful as molecular markers for
 CC studying inheritance and multigene traits in a plant breeding program.
 CC The cdps are useful for producing purified corn-ear polypeptides by
 CC recombinant techniques. They are also useful in diagnostic assays to
 CC detect or confirm conditions or diseases associated with abnormal levels
 CC of cdp expression. ABX81641-ABX89140 represent corn ear-derived
 CC polynucleotides (cdps) of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from the USPTO web site at
 CC seqdata.uspto.gov/psipds/identity.html
 CC
 SO Sequence 272 BP; 79 A; 52 C; 68 G; 67 T; 0 U; 6 Other;

	Query Match	4.0%	Score 75; DB 7; Length 272;
	Best Local Similarity	57.6%	Pred. No. 8.7e-11;
	Matches 148; Conservative	0; Mismatches 106; Indels 3; Gaps 1;	
Oy	1009 TATGAGGCTTTCAGACGCCATTCCATGGTGCAAGGC CATCGTAGATGA GTGT		1068
Dd	2 TATGAAACAAGTAAGA CACTCATCCTAATGTTCCCTCGGGCACCGTGATGA GCCTN		61
Oy	1069 CTGATGTTTTGAGAGCA GTGCCACTGSGCTA TTITGGAGGCCG AACGCTTCCACC GGAGTTA		1128
Dd	62 TTAATAATTGAAAAGCT CACTGTGGGCTAAC ATGMAAGCTGA ACGCTGCATAA ACACCTT		121
Oy	1129 GGTGAGATGGGGTT AGATGAATGGCCGG---GGTCAGAA CGCAGATGTTT TCCTGGA		1185
Dd	122 GTTAANCAA GSTPACAGACAGGAATTCAG GAACCTCA GCCANNGGTTTGC ATTGTGCTGGT		181
Oy	1186 GGTTGTCGCCAAC TGATGGCTTCCCTTGCA ACAAGACAGATCTG SACAATTCATCAA		1245
Dd	182 GGGAAAAAGCNCCTTAT NGTTCTTCTACCA AAACAAGATATGAGCATT TAACAAG		241
Oy	1246 CACTCTCAGGCAAAAC	1262	
Dd	242 CATGCCATGATAAAGC	258	
RESULT 4			
ABX82003			
ID	ABX82003 standard; cDNA; 274 BP.		
AX			
XX	ABX82003;		
DT	24-APR-2003 (first entry)		
DE	Corn ear-derived polynucleotide (cpd) #463.		
KW	Corn ear-derived polynucleotide; cpd; cDNA library; SATMON022; SATMON023; structural gene; functional gene, regulatory gene, corn ear-specific profile; gene transcription; gene expression; hybrid plant; desirable trait expression; plant breeding program; inheritance; desired characteristic; growth; development; disease resistance; environmental adaptability; quality; yield; multigene trait; plant; gene; ss.		
XZ	Zea mays.		
OS			
PV	US6476212-B1.		
PN			
XX			
PD	05-NOV-2002.		
XX			
FY	14-MAY-1999; 99US-00313294.		
PR			
XX	26-MAY-1998; 98US-0086722P.		
PA	(INCY-) INCYTE GENOMICS INC.		
P1	Laljudi RV, Ito LY, Sherman BK;		
P1	WIPI; 2003-208840/20.		
DR			
PT			
XX			
PS			
XX	Example; SEQ ID NO 463; 390bp; English.		
CC	The present invention relates to the isolation of corn ear-derived polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022 and SATMON023. Some of the cdps uniquely identify structural, functional, and regulatory genes of corn ear. The polynucleotides sequences are useful for detecting cdps in a sample, for producing a corn ear-specific profile of gene transcription, for detecting altered gene expression in indured or hybrid plants, and for screening several molecules for specific		

Query Match	Best Local Similarity	Score	DB	Length
Matches 156; Conservative 0; Mismatches 98; Indels 4; Gaps 2;	4.08; 60.5%; 1.3e-10;	74.4; 7; 1.3e-10;	DB 7; 1.3e-10;	274;
Query	1009	TATGAGGCTCTTAAAGAGCCGATTCCTATGCT--CCACAGAGGCCATGTGGAGTAGCT	1067	
Db	2	TATGAAACCAAGTAAAGACGCTCATGCTTATGCTCTCTGCGGACCGTGNATGACGCT	61	
Qy	1068	TCTGATGTTTGAAGACAGCTGCACTGCTATTTTGAAGGCCGACGCTCCACCGGAGTT	1127	
Db	62	GTTTATATTTTGAAGAGCTCAGCTGTGGGCTCTACAGAGCTGAAAGCTTGCATTAACACTT	121	
Qy	1128	AGCTGAGATGGGGTTAGATGAAATTTGCTGG--GGTCAGAAAGCAGATATGTTTCTGG	1184	
Db	122	TGTTATATCAAGGTATCAGACAGAAATTCATGCGACCTACGAAAGTTGATTTGGCTGG	181	
Qy	1185	AGGTGTTCCGCAACTGATGATGCTTCTTTCGACAGACGAAAGATGTGACATATTCATCA	1244	
Db	182	TGGGAAAGGCAACTATATGTTCTTTCGCAACCAAGAGATATGAGGAGCATTTAAACA	241	
Qy	1245	ACACTCTCAAGGCAAAAC	1262	
Db	242	GCAATGCCAGGTAAAGC	259	
RESULT 5	ABL71922	ABL71922	ABL71922	ABL71922
ID	ABL71922	standard; cDNA; 310 BP.		
AC	ABL71922;			
XX	14-MAY-2002	(first entry)		
DE	Corn tassell-derived polynucleotide (cdps) SEQ ID NO:1296.			
XX	Corn; corn tassell-derived polynucleotide; cdps; hybrid breeding; CDPS;			
KW	inheretance; characterisic; growth; development; disease resistance;			
KW	environmental adaptability; quality; yield; molecular marker;			
KW	multigene trait; plant breeding; corn tassell; gene; ss.			
XX	Zea mays.			
OS	US2001051335-A1.			
XX	13-DEC-2001.			
PD	16-APR-1999;	99US-00294093.		
XX	21-APR-1998;	98US-0082567P.		
XX	(LATG/) LATGUDI R V.			
PA	(ITOL/) ITO L Y.			
PA	(SHER/) SHERMAN B K.			

QY 1278 GTTGAATCATACCAAGAGATGTTGTTAAAGAGCTGAGGACAGATCTTGAGGACAAATCA 1337
 DB 1860 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 1919
 QY 1338 GCAAGCTGAATCTCTTTAAGAACAGCTCTCAAAAACAAACAGCCAGCCAGGCTGTTGA 1397
 DB 1920 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGATGAGCA 1979
 QY 1398 GGAATCTCTGAAATTTATGAGGAGAGAGCTGGGTGAGAACTGACAGGATTAATCGATCGT 1457
 DB 1980 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 2039
 QY 1458 GAGACAGAGAACTAAGATGACAGATGAAACAGAAACAGGAAAGATGATGACACAGCAG 1517
 DB 2040 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 2099
 QY 1518 GTTTTTCATGATTCATCAAAACAGATTCATGAAAGAGACGCAAGAGAGAGAAATTT 1577
 DB 2100 GCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGCA 2159
 QY 1578 CGAGATGTTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 1637
 DB 2160 GATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGATGA 2219
 QY 1638 TCCCTCTAGCAATGACGATTCGCGAAAGAGAGCTGAGAGAGTGTCAAGCTTCATCGAGTT 1697
 DB 2220 GCAGCAGCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGCAGCAGCAGCAGAGAGAGAGCA 2279
 QY 1698 TCAAGAGAAAGATGAGAGAGCTTTGTGGAAGAGAGGAGATGCTGATTAAGATCAAGA 1757
 DB 2280 GCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGCAGATCAAGA 2339
 QY 1758 GAAAGAAATGGAACATGAAGAAGAGCATCAAGAGAGATTTGATCTGAGAGAAAGA 1817
 DB 2340 GCAGAGATTTAGAGAGCAGCAGCAGAGATTTAGAGAGAGCAGAGAGAGATTTAGAGAGCA 2399
 QY 1818 ATTGATGAGCTTTGGAACAG 1839
 DB 2400 GCAGCAGAGATTTAGAGAGCAG 2421

RESULT 7
 AAF82901
 ID AAF82901 standard; DNA; 3489 BP.

AC AAF82901;
 DT 11-SEP-2003 (revised)
 DT 06-AUG-2003 (revised)
 DT 29-JUN-2001 (first entry)
 DE Nucleotide sequence of KSHV tethering protein, LANA.
 XX Histone H1, tethering protein; LANA; gene therapy; multiple sclerosis;
 KM Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
 KW KSHV; latency-associated nuclear antigen; LANA; ds.
 XX
 OS Human herpesvirus 8.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..3489
 FT /*tag= a
 XX
 FN WO200125484-A2.
 XX
 PD 12-APR-2001.
 XX
 PF 29-SEP-2000; 2000WO-US026908.
 XX
 PR 01-OCT-1999; 99US-00410399.
 XX
 PA (UNMI) UNIV MICHIGAN.

XX
 PI Robertson ES, Colter MA;
 XX
 DR WPI: 2001-281736/29.
 DR P-PSDB; AAB62331.
 XX
 PT A composition for use in gene therapy comprises an expression vector that
 PT includes a nucleic acid sequence encoding a nucleic acid binding protein.
 XX
 PS Disclosure; Fig 9A; 60pp; English.
 XX
 CC The invention provides a composition comprising nucleic acid, histone H1
 CC protein and expression vector operably encoding a protein suitable
 CC for tethering the nucleic acid to the histone H1 protein, where the
 CC tethering protein is LANA. The composition is useful in aiding the
 CC retention of the viral DNA in the host cell. The viral vector encodes a
 CC protein suitable for tethering DNA to Histone H1. Methods for screening
 CC for compounds which are agonistic or antagonistic for the tethering of
 CC viral proteins to histone H1 and DNA binding sites are useful for
 CC developing the method of viral transfer. The composition has applications
 CC to gene therapy, including the treatment of multiple sclerosis,
 CC Parkinson's disease, Huntington disease and diabetes. The present
 CC sequence represents the nucleotide sequence of the Kapost's sarcoma
 CC associated herpesvirus (human herpesvirus 8) latency-associated nuclear
 CC antigen (LANA), which acts as a tethering protein. (Updated on 06-AUG-
 CC 2003 to correct OS field.) (Updated on 11-SEP-2003 to standardise OS
 CC field)
 SQ Sequence 3489 BP; 1053 A; 862 C; 1137 G; 437 T; 0 U; 0 Other;
 Query Match 3.1%; Score 58.8; DB 4; Length 3489;
 Best Local Similarity 43.4%; Pred. No. 1.4e-05;
 Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;
 QY 1218 GAAGCAAGATTCGACATATTCATCAACACTCTTAAGCAAAACAGGCTGAATTGCA 1277
 DB 1800 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCA 1859
 QY 1278 GTTGAATCATACCAAGAGATGTTGTTAAAGAGCTGAGGACAGATCTTGAGGACAAATCA 1337
 DB 1860 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCA 1919
 QY 1338 GCAAGCTGAATCTCTTTAAGAACCAAGCTCTCAAAAACAAACAGCCAGGCTGTTGA 1397
 DB 1920 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 1979
 QY 1398 GGAATCTCTGAAATTTATGAGGAGAGAGCTGGGTGAGAACTGACAGGATTAATCGATCGT 1457
 DB 1980 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 2039
 QY 1458 GAGACAGAGAACTAAGATGACAGATGAAACAGAAACAGGAAAGATGATGACACAGCAG 1517
 DB 2040 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCA 2099
 QY 1518 GTTTTTCATGATTCATCAAAACAGATTCATGAAAGAGACGCAAGAGAGAGAAATTT 1577
 DB 2100 GCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGATGAGCAGCAGCAGAGATGAGCA 2159
 QY 1578 CGAGATGTTGCAAGCAGCAGAACTGCGCAAGGTTTGGCCACAGCAGCAGCAACATTAA 1637
 DB 2160 GATGATGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGATGA 2219
 QY 1638 TCCCTCTAGCAATGACGATTCGCGAAAGAGAGCTGAGAGAGTGTCAAGCTTCATCGAGTT 1697
 DB 2220 GCAGCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGCAGCAGCAGCAGAGATGAGCA 2279
 QY 1698 TCAAGAGAAAGATGAGAGAGCTTTGTGGAAGAGAGGAGATGCTGATTAAGATCAAGA 1757
 DB 2280 GCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCAGCAGAGATGAGCA 2339
 QY 1758 GAAAGAAATGGAACATGAAGAAGAGCATCAAGAGAGATTTGATCTGAGAGAAAGA 1817
 DB 2340 GCAGAGATTTAGAGAGCAGCAGCAGAGATTTAGAGAGAGCAGAGAGATTTAGAGAGCA 2399

DR MPI, 1999-069741/06.
 XX Kaposi's sarcoma-associated herpes virus nucleic acid - encodes
 PT dihydrofolate reductase and is useful for treatment, prophylaxis or
 PT diagnosis of Kaposi's sarcoma.
 XX
 PS Disclosure, Col 155-182, 109pp; English.
 CC This sequence is a fragment of the Kaposi's sarcoma-associated
 CC herpesvirus (KSHV) LVR (long unique region). This fragment contains
 CC coding regions for ORF65 which encodes capsid protein IV, ORF66, ORF67
 CC which encodes tegument protein IV, ORF68 which encodes a glycoprotein,
 CC ORF69, K12 which encodes Kaposin, K13, ORF72 which encodes cyclin D,
 CC ORF73 which encodes immediate early protein (IEP), K14 which encodes
 CC (v-adh), ORF74 which encodes G-protein coupled receptor, ORF75 which
 CC encodes tegument protein/FGFARF, K15. KSHV is a new human Herpesvirus
 CC (HHV8) believed to cause Kaposi's sarcoma (KS) which is the most common
 CC form of neoplasm occurring in persons with acquired immune deficiency
 CC syndrome (AIDS). The DHFR protein is useful for vaccination, prophylaxis,
 CC diagnosis and treatment of a subject with Kaposi's sarcoma and for
 CC detecting expression of a DNA virus associated with Kaposi's sarcoma in a
 CC cell. (Updated on 17-OCT-2003 to standardise OS field)
 XX
 SQ Sequence 32207 BP, 7229 A, 9156 C, 8713 G, 7109 T, 0 U, 0 Other;
 Query Match 3.1%; Score 58.8; DB 2; Length 32207;
 Best Local Similarity 43.4%; Pred. No. 4.5e-05;
 Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;
 QY 1218 GAAGCAAGATCTGGACATATTCATCACTCTCAAGCAAAACAGGCTGAATTCGA 1277
 DB 20197 GCAGCGAGATGAGCAGACAGAGATGAGCAGCAGATGAGCAGCAGAGATGAGCA 20138
 QY 1278 GTTGAATATATCCAGAGATGTTGTAAGAGCTGAGCAGATCTGAGACATCA 1337
 DB 20137 GCAGCGAGATGAGCAGACAGAGATGAGCAGCAGATGAGCAGCAGAGATGAGCA 20078
 QY 1338 GCAGCTGAATCTTAAAGAACTCTCAAAACAGCAAGCAAGGCTGCTTGA 1397
 DB 20077 GCAGCGAGATGAGCAGACAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCA 20018
 QY 1398 GGAATCTCTGGAATATTATGAGCGAAGCTGCTAGAACTGCAGAGATTAATCGATCGT 1457
 DB 20017 GCAGCGAGATGAGCAGACAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCA 19958
 QY 1458 GAGACAGAACTAAGATGAGATGAAAGAAACAGGAAAGATGATGACACACAG 1517
 DB 19957 GCAGCGAGATGAGCAGACAGAGATGAGCAGCAGAGATGAGCAGCAGAGATGAGCA 19898
 QY 1518 GTTTTTCATGATTTCAATCAATCAGATTCATGAAAGAAAGACGAAAGAGAGAAATTT 1577
 DB 19897 GCAGCGAGAGATGAGCAGACAGAGATGAGCAGAGATGAGCAGAGATGAGCAGACGA 19838
 QY 1578 CGAGATGTTGACAGCAGACAGAAAGTCCAAAGTTGTTGCGCAGCAGCAGACAGAACTTAA 1637
 DB 19837 GGATGAGCAGCAGCAGCAGATGAGCAGCAGCAGAGATGAGCAGCAGCAGAGATGAGCA 19778
 QY 1638 TTCCTTACATGAGATGATTCGCGAAAGAGAGCTGAGAAAGTCAAGCTTATCGAGTT 1697
 DB 19777 GCAGCGAGCAGAGATGAGCAGACAGAGATGAGTGAACAGGAGCAGCAGAGAGACAGA 19718
 QY 1698 TCAGAGAAAGAGATGAGAGAGTTTGTGAAGAGAGAGAGATGCTGATAAAGATCAAGA 1757
 DB 19717 GCAGCGAGAGAGCAGAGAGAGAGAGATTGAAGAGAGAGAGAGAGATTGAAGATGAGA 19658
 QY 1758 GAAGAAGATGAAGAAGATGAAGAAGAGATCAAGAGAGATATTGATCTGAGAAAGA 1817
 DB 19657 GCAGAGATTAGAGAGCAGAGAGAGAGATTAGAGAGAGAGAGAGAGATTAGAGAGAGA 19598
 QY 1818 ATTTGATGAGGCTTTGGAACAG 1839
 DB 19597 GCAGCGAGATTAGAGAGAGAG 19576

RESULT 10
 AAV19941/c
 ID AAV19941 standard; DNA; 137507 BP.
 XX
 AC AAV19941;
 XX
 DT 27-AUG-2003 (revised)
 DT 03-AUG-1998 (first entry)
 XX
 DE KSHV long unique coding region and terminal repeat.
 XX
 XX KSHV, HHV8; human herpes virus 8; macrophage inflammatory protein II;
 XX interleukin-6; IL-6; interferon regulatory factor; rheumatoid arthritis;
 XX complement-binding protein; glycoprotein; capsid protein IV; infection;
 XX immediate early protein; Kaposi's sarcoma; protective vaccine; lymphoma;
 XX lymphoproliferative disease; leukaemia; splenomegaly; mycosis fungoides;
 XX HIV immune status; anti-inflammatory agent; therapy; ds.
 XX
 OS Human herpesvirus 8.
 XX
 FH Key Location/Qualifiers
 FT CDS 1142..2794
 FT /tag= a
 FT /product= "complement-binding protein"
 FT 8659..11236
 FT /tag= b
 FT /product= "glycoprotein B"
 FT complement(17261..17875)
 FT /tag= c
 FT /product= "interleukin 6"
 FT complement(21548..21832)
 FT /tag= d
 FT /product= "macrophage inflammatory protein II"
 FT complement(27137..27424)
 FT /tag= e
 FT /product= "interferon regulatory factor 1"
 FT 28661..29741
 FT /tag= f
 FT /product= "protein T1.1"
 FT complement(58976..60175)
 FT /tag= g
 FT /product= "glycoprotein M"
 FT complement(69412..69915)
 FT /tag= h
 FT /product= "glycoprotein L"
 FT complement(88410..88910)
 FT /tag= i
 FT /product= "interferon regulatory factor 2"
 FT 89600..90541
 FT /tag= j
 FT /product= "interferon regulatory factor 3"
 FT 90173..90643
 FT /tag= k
 FT /product= "glycoprotein X"
 FT complement(93636..94127)
 FT /tag= l
 FT /product= "interferon regulatory factor 4"
 FT complement(111931..112443)
 FT /tag= m
 FT /product= "capsid protein IV"
 FT complement(123808..127296)
 FT /tag= n
 FT /product= "immediate early protein"
 XX
 PN WO9804576-A1.
 XX
 PD 05-FEB-1998.
 XX
 PF 22-JUL-1997; 97WO-US013346.
 XX
 XX 25-JUL-1996; 96US-00686243.
 PR 25-JUL-1996; 96US-00686349.

PR 25-JUL-1996: 96US-00686350.
 PR 25-JUL-1996: 96US-00687253.
 PR 05-SEP-1996: 96US-00708678.
 PR 10-OCT-1996: 96US-00728323.
 PR 13-NOV-1996: 96US-00747887.
 PR 13-NOV-1996: 96US-00748640.
 PR 29-NOV-1996: 96US-00757669.
 (UYCO) UNIV COLUMBIA NEW YORK.
 PA Chang Y, Bohenzky RA, Russo JJ, Edelman IS, Moore PS;
 PI WPI; 1996-130615/12.
 DR New nucleic acid encoding Kaposi's sarcoma associated herpes virus
 PT proteins - useful for, e.g. detecting levels of HHV8 in, and preparation
 PT of vaccines for treatment of, HIV patients.
 PS Example 2; Page 135-203; 230pp; English.
 XX This sequence represents the long unique region and terminal repeat of
 CC the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
 CC as human herpes virus 8 (HHV8). This sequence selected from: (a) viral
 CC invention which encode KSHV polypeptides selected from: (a) viral
 CC macropage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
 CC (c) viral RFP 1; (d) complement-binding protein; and (e) immediate early protein
 CC encoded by ORF73. Labeled probes for the nucleic acid, proteins encoded
 CC by it, and antibodies (Ab) specifically for diagnosis of Kaposi's sarcoma, in body
 CC fluids or tissue samples. HHV8 infections can be treated with antiseise
 CC or triplex forming molecules or agents that bind specifically to the
 CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
 CC while the protein can be used in protective vaccines. Ab may also be used
 CC to differentiate between lymphomas, and HHV8 may be implicated in many
 CC other lymphoproliferative diseases such as lymphomas, leukemia,
 CC CC splenomegaly and mycosis fungoides. Cells and animals containing the
 CC nucleic acid are useful for drug screening. HHV8-derived peptidase gene
 CC can be used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
 CC can be inhibited with methotrexate. These can also be used to determine
 CC the immune status of a patient infected with HIV. HHV8 derived protein
 CC viral MIP III may be used as an anti-inflammatory agent for, e.g.
 CC treating rheumatoid arthritis. This sequence is stated as containing 81
 CC open reading frames. (Updated on 27-AUG-2003 to correct OS field.)
 SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 U; 0 Other;
 Query Match 3.1%; Score 58.8; DB 2; Length 137507;
 Best Local Similarity 43.4%; Pred. No. 9.7e-05;
 Matches 270; Conservative 0; Mismatches 352; Indels 0; Gaps 0;
 QY 1218 GAGCAAGATCTGCACATATTCATCAACTCTCAAGGCAAAACAGCTGTAATTCGA 1277
 DB 125497 GCAGCAGATGAGCGACGACGATGAGCAGCAGATGAGCAGCAGATGAGCA 125438
 QY 1278 GTTGAATCATCAAGAGATGTTTAAAGACTGAGCAGATCTCTGAGACATCA 1337
 DB 125437 GCAGCAGATGAGCGACGACGATGAGCAGCAGATGAGCAGCAGATGAGCA 125378
 QY 1338 GCACTGAACTACTTTAAGACACAGCTCTCAAAACAGACAGCAGCAAGCTGCTTGA 1397
 DB 125377 GCAGCAGATGAGCGACGACGATGAGCAGCAGATGAGCAGCAGATGAGCA 125318
 QY 1398 GGAATCTGGAATTAATGAGCAGAGCTGCTGAAACTGAGAGGATATCGATCGT 1457
 DB 125317 GCAGCAGATGAGCAGCAGAGATGAGCAGCAGATGAGCAGCAGATGAGCA 125258
 QY 1458 GAGCAGAGAACTTAAGTGCAGATGAACAGACAGGAGATGATGACACAGCAG 1517
 DB 125257 GCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCA 125198
 QY 1518 GTTTTATGATTCATCAATCAACAGATCCATGAAAGAGAGCGCAAGAGATTT 1577

DB 125197 GAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCAGATGAGCAGCA 125138
 QY 1578 CGAATGTTGACAGCAGCAGAACTGCCAGGTTTGGCCAGCAGCAGCAATTTAA 1637
 DB 125137 GATGATGACAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGATGAGCA 125078
 QY 1638 TCCCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1697
 DB 125077 GCAGCAGCAGATGAGCAGCAGCAGCAGATGAGCAGCAGATGAGCAGCAGCA 125018
 QY 1698 TCAAGAGAAAGATGAGAGGATTTGTGAAAGAGAGAGATGCTGATTAAGATCAAGA 1757
 DB 125017 GCAGCAGCAGCAGCAGCAGCAGATGAGCAGCAGCAGATGAGCAGCAGCA 124958
 QY 1758 GAAAGATGAGAGCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1817
 DB 124957 GCAGCAGATGAG 124898
 QY 1818 ATTTGATGAGGCTTGAACAG 1839
 DB 124897 GAGCAGCAGATGAGAGAGAGAG 124876
 RESULT 11
 ID AB199537 standard; cDNA; 1080 BP.
 XX AB199537;
 XX 07-MAR-2002 (first entry)
 DE Mouse ischaemic condition related cDNA sequence SEQ ID NO:537.
 XX Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 XX vasospastic ischaemia; ischaemic condition; ischaemic disease; ss.
 OS Mus musculus.
 PN WO20018188-A2.
 XX 22-NOV-2001.
 PD 18-MAY-2001; 2001WO-0P004192.
 PF 18-MAY-2000; 2000JP-00145977.
 PR 18-MAY-2000; 2000JP-00145977.
 PA (YUNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.
 XX Yuni K, Arai S, Takahashi Y, Nagata T, Ishii Y;
 PI Ishikawa K, Arai S, Takahashi Y, Nagata T, Ishii Y;
 DR WPI; 2002-034733/04.
 DR P-PSDB; ABB57221.
 XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring
 PT expression levels of particular genes defined in the specification or by
 PT determining the expression profile of a gene group comprising these
 PT genes.
 PS Claim 2; Page 1472-1473; 2690pp; English.
 XX The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (I) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (I). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasospastic ischaemia) by measuring the
 CC expression levels of particular genes (AB199202 to AB199912, encoding the
 CC expression sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The expression
 CC levels or expression profiles for ischaemic condition-improving drugs or
 CC indicator when screening for ischaemic condition-improving drugs or
 CC therapeutics for ischaemic diseases. AB199913 and AB199914 represent
 CC primers for a mouse ischaemic condition related sequence, which are used

CC in the exemplification of the present invention
XX Sequence 1080 BP; 370 A; 191 C; 390 G; 129 T; 0 U; 0 Other;

Query Match 3.0%; Score 56.2; DB 6; Length 1080;
Best Local Similarity 48.0%; Pred. No. 4.1e-05;
Matches 160; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

QY 1466 GAACTAAGATGCAATGACACAGACGAGGAGATGATGACACGACAGGTTTTTCA 1525
DB 701 GAAAAAGAAAG 160
QY 1526 TGGATTCATCAACAGATCCATGAAAGAGAGAGAGAGAGAGAGATTTTCAAGATG 1585
DB 761 AGGAG 820
QY 1586 TGCAGACAG 1645
DB 821 AGAAG 880
QY 1646 GCAATGACGATTGCCGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1705
DB 881 AGAAG 940
QY 1706 AAGAGATGAG 1765
DB 941 AGAAG 1000
QY 1766 TGGAGACATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1798
DB 1001 AGAAG 1033

RESULT 12

AAQ87587
ID AAQ87587 standard; DNA; 1686 BP.

AC AAQ87587;

DT 27-AUG-2003 (revised)
DT 19-DEC-1995 (first entry)

DE DNA encoding Leucocytozoan protozoa structural protein epitope.

KM leucocytozoan protozoa; structural protein; epitope; vaccine; fowl;

KM leucocytozoanosis; treatment; ss.

OS Leucocytozoan.

PN JP07089995-A.

PD 04-APR-1995.

PF 10-SEP-1993; 93JP-00226078.

PR 10-SEP-1993; 93JP-00226078.

PA (DOB-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.

PA (NISI-) NISEIKEN KK.

DR WPI; 1995-167252/22.

DR P-PSDB; AAR70491.

PT Immune inducing polypeptide against Leucocytozoan protozoa - useful in
production of vaccines for treatment of Leucocytozoanosis in fowl.

PS Claim 1; Page 12-14; 20p; Japanese.

CC AAQ87587-89 encode polypeptides having a whole or partial epitope of a
structural protein of Leucocytozoan protozoa (see AAR70491-93). The
polypeptides and DNA encoding them are useful in the production of
vaccines for the treatment of Leucocytozoanosis of fowl. (Updated on 27-
AUG-2003 to correct OS field.)

XX Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 U; 0 Other;

Query Match 2.9%; Score 55; DB 2; Length 1686;
Best Local Similarity 44.4%; Pred. No. 0.00011;
Matches 220; Conservative 0; Mismatches 275; Indels 0; Gaps 0;

QY 1305 AAGAGAGTGAAGGAGATCTTGAAGACATGACAGCTGAATCTTTAAGACAGCT 1364
DB 88 AATGTGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 147
QY 1365 TTCAAAACAGAAACAGCAGCAGAGTCTTGAAGATCTTGAAGAAATTAAGCAGAA 1424
DB 148 CAAGAAACAAGAAACAAGAAACAAGAAATCTTAAGAAACAAGAAACAAGAAAGAA 207
QY 1425 GCTGCTAGAACTGCAAGAGATTAATCGATTCGTGAGACAGAGAACTTAAGATCAGATTA 1484
DB 208 CAAGAAAGAGAAAGATGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 267
QY 1485 ACAGAAACAGGAGAGATGATGACACAGACAGGTTTTCATGATTCATCAACAGAT 1544
DB 268 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 327
QY 1545 CCATGAAG 1604
DB 328 GAAACAAGATGAAGATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 387
QY 1605 CAAAGTTGTTGCCAGCAGCAGCAGAACTTAATCCCTTACCAATGAAGATTTCCGAAA 1664
DB 388 AATGAAGAGAGAAACAAATGAAGATGAAGAAAGAGAGAGAGAGAGAGAGAGAGAA 447
QY 1665 GAGAGCTGAGAGAGTCAAGCTTCATCGAGTTTCAAGAGAAAGAGATGAGAGAGTTTGT 1724
DB 448 GAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 507
QY 1725 GGAAG 1784
DB 508 GATGAAG 567
QY 1785 GCATCAGAGAGAT 1799
DB 568 GCTGAAAAAGAAAT 582

RESULT 13

AAT91902
ID AAT91902 standard; DNA; 4000 BP.

AC AAT91902;

DT 07-MAY-1998 (first entry)

DE Mannose-1-phosphate transferase gene MNNA.

KM Mannose-1-phosphate transfer; MNNA gene; enzyme; yeast; regulatory gene;

KM human; high mannose type neutral saccharide chain; ss.

OS Saccharomyces cerevisiae.

FT Key Location/Qualifiers

FT CDS 418..3953

FT /tag= a

FT /product= "MNNA"

PN JP09266792-A.

PD 14-OCT-1997.

PF 29-MAR-1996; 96JP-00075667.

PR 29-MAR-1996; 96JP-00075667.

PA (AGEN) AGENCY OF IND SCI & TECHNOLOGY.

XX
DR WPI; 1997-553460/51.
DR P-PDB; AAM30763.
XX
PT Positive regulatory gene of mannose-1-phosphate transfer in yeast
XX useful for high mannose type neutral saccharide chain production.
XX
PS Claim 2; Page 14-17; 23pp; Japanese.
XX
XX This sequence represents the gene of the invention, designated MN4,
CC encoding a protein which positively regulates mannose-1-phosphate
CC transfer in yeast. The gene is useful for the preparation of human high
CC mannose type neutral saccharide chain
SQ Sequence 4000 BP; 1453 A; 750 C; 770 G; 1027 T; 0 U; 0 Other;
2.9%; Score 55; DB 2; Length 4000;
Query Match Best Local Similarity 47.9%; Pred. No. 0.00018;
Matches 191; Conservative 0; Mismatches 205; Indels 3; Gaps 1;
QY 1441 GAGGATATCGATCGTGAACAGAGAACTTAAGTCAAGTGAAGACAGAAACAGGAGAG 1500
DB 3539 GAG 3598
QY 1501 ATGATGACACGACGACGATTTTCATGATTCATCAATCAACAGATCCATGAAGAGAC 1560
DB 3599 AAGGAG 3655
QY 1561 GCAAG 1620
DB 3656 GAG 3715
QY 1621 CAGCGCAGAACTTAATCCCTTAGCAATGACATTCGCCAAGAGAGAGAGAGAG 1680
DB 3716 AAG 3775
QY 1681 TCAGAGCTTCAGAGTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 3776 GAG 3835
QY 1741 CTGATTAAGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 1800
DB 3836 AAGCAG 3895
QY 1801 TTGATCTGAG 1839
DB 3896 GAAAG 3934
RESULT 14
AAC40071
ID AAC40071 standard; DNA; 1140 BP.
XX
AC AAC40071;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 26940.
DE
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway; metabolic pathway;
KM promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
OS
PN EP1033405-A2.
PN
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130442P.
PR 23-APR-1999; 99US-0130510P.
PR 28-APR-1999; 99US-0130891P.
PR 30-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 30-APR-1999; 99US-0132407P.
PR 04-MAY-1999; 99US-0132484P.
PR 05-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134222P.
PR 18-MAY-1999; 99US-0134370P.
PR 19-MAY-1999; 99US-0134768P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136332P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 17-JUN-1999; 99US-0139492P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140655P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 09-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142803P.
PR 12-JUL-1999; 99US-0142920P.
PR 13-JUL-1999; 99US-0142977P.
PR 14-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.

PR	15-JUL-1999	99US-01440051
PR	16-JUL-1999	99US-01440052
PR	16-JUL-1999	99US-01440066
PR	19-JUL-1999	99US-01443357
PR	19-JUL-1999	99US-01443311
PR	19-JUL-1999	99US-01443332
PR	19-JUL-1999	99US-01443333
PR	19-JUL-1999	99US-01443344
PR	19-JUL-1999	99US-01443355
PR	20-JUL-1999	99US-01443352
PR	20-JUL-1999	99US-01446322
PR	20-JUL-1999	99US-01446824
PR	21-JUL-1999	99US-01446814
PR	21-JUL-1999	99US-01450066
PR	21-JUL-1999	99US-01450088
PR	22-JUL-1999	99US-01450087
PR	22-JUL-1999	99US-01450872
PR	22-JUL-1999	99US-01450892
PR	22-JUL-1999	99US-01451152
PR	23-JUL-1999	99US-01451155
PR	23-JUL-1999	99US-01452188
PR	23-JUL-1999	99US-01452242
PR	26-JUL-1999	99US-01452766
PR	27-JUL-1999	99US-01455133
PR	27-JUL-1999	99US-01455187
PR	27-JUL-1999	99US-01459187
PR	28-JUL-1999	99US-01459119
PR	28-JUL-1999	99US-01459511
PR	02-AUG-1999	99US-01463366
PR	02-AUG-1999	99US-01463388
PR	02-AUG-1999	99US-01463892
PR	03-AUG-1999	99US-01470382
PR	04-AUG-1999	99US-01472044
PR	04-AUG-1999	99US-01473022
PR	05-AUG-1999	99US-01473922
PR	06-AUG-1999	99US-01472606
PR	06-AUG-1999	99US-01473032
PR	06-AUG-1999	99US-01474316
PR	09-AUG-1999	99US-01474932
PR	09-AUG-1999	99US-01474935
PR	10-AUG-1999	99US-01481719
PR	11-AUG-1999	99US-01483131
PR	12-AUG-1999	99US-01483112
PR	13-AUG-1999	99US-01485652
PR	13-AUG-1999	99US-01486842
PR	16-AUG-1999	99US-01493662
PR	16-AUG-1999	99US-01491752
PR	16-AUG-1999	99US-01494262
PR	20-AUG-1999	99US-01497222
PR	20-AUG-1999	99US-01497322
PR	20-AUG-1999	99US-01499292
PR	23-AUG-1999	99US-01499022
PR	23-AUG-1999	99US-01499302
PR	25-AUG-1999	99US-01505662
PR	26-AUG-1999	99US-01508842
PR	27-AUG-1999	99US-01510652
PR	27-AUG-1999	99US-01510662
PR	27-AUG-1999	99US-01510802
PR	30-AUG-1999	99US-01513032
PR	31-AUG-1999	99US-01514382
PR	01-SEP-1999	99US-01519302
PR	07-SEP-1999	99US-01523632
PR	10-SEP-1999	99US-01530702
PR	13-SEP-1999	99US-01537582
PR	15-SEP-1999	99US-01540182
PR	16-SEP-1999	99US-01540192
PR	20-SEP-1999	99US-01547792
PR	22-SEP-1999	99US-01551392
PR	23-SEP-1999	99US-01554862
PR	24-SEP-1999	99US-01555692
PR	28-SEP-1999	99US-01564582
PR	29-SEP-1999	99US-01565362
PR	05-OCT-1999	99US-01571732
PR	05-OCT-1999	99US-01577532

PR	06-OCT-1999;	99US-0157665P
PR	07-OCT-1999;	99US-0158029P
PR	08-OCT-1999;	99US-0158232P
PR	12-OCT-1999;	99US-0158369P
PR	13-OCT-1999;	99US-0158233P
PR	13-OCT-1999;	99US-0159284P
PR	13-OCT-1999;	99US-0159255P
PR	14-OCT-1999;	99US-0159359P
PR	14-OCT-1999;	99US-0159310P
PR	14-OCT-1999;	99US-0159311P
PR	14-OCT-1999;	99US-0159637P
PR	14-OCT-1999;	99US-0159638P
PR	18-OCT-1999;	99US-0159584P
PR	21-OCT-1999;	99US-0160741P
PR	21-OCT-1999;	99US-0160767P
PR	21-OCT-1999;	99US-0160768P
PR	21-OCT-1999;	99US-0160770P
PR	21-OCT-1999;	99US-0160814P
PR	21-OCT-1999;	99US-0160815P
PR	22-OCT-1999;	99US-0160960P
PR	22-OCT-1999;	99US-0160981P
PR	22-OCT-1999;	99US-0160989P
PR	25-OCT-1999;	99US-0161405P
PR	25-OCT-1999;	99US-0161406P
PR	26-OCT-1999;	99US-0161359P
PR	26-OCT-1999;	99US-0161360P
PR	26-OCT-1999;	99US-0161361P
PR	28-OCT-1999;	99US-0161970P
PR	28-OCT-1999;	99US-0161982P
PR	28-OCT-1999;	99US-0162142P
PR	29-OCT-1999;	99US-0162193P

Query Match	2.8%	Score 53.2;	DB 3;	Length 1140;
Best Local Similarity	45.2%;	Pred. No. 0.0003;		
Matches 196;	Conservative 0;	Mismatches 238;	Indels 0;	Gaps 0;

QY	1381	AAGCCAAAGTGCCTTGAAGAACTCTCTGGAAATTATAGCCAGAAAGTCGCCGTGAACCTGCA	1440
Db	473	CACAGGATGATGGTGAAGATGGTCCACAAAAACATAGTGGTAAGATTATGATGATGGA	532
QY	1441	GAGGATATATCGATCGTGAAGACAGAGAAGCTAAGATGACAGATGAACAGAAACAGGAAGAG	1500
Db	533	GATGAGAAAGTAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG	592
QY	1501	ATGATGACACGACAGAGGTTTTTCATGATTCAAACAGATCCATGAAGAGAGAC	1560
Db	593	AACCTGAAGTGTGATGATTTAAGAAAGAGAGAGAGAGAGAGAGAGAGAGATCAT	652
QY	1561	GCAAGGAGGAGATTTGAGATGTTGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1620
Db	653	GATATGATGATTTATGATGAG	712
QY	1621	CAGCAGCAGAACATTATCCCTCTGACATGACGATTTCCGAAAGAGAGAGAGAGAGAGT	1680
Db	713	AAGAAAGAGAGAGAGAGATTTAATGATGATGATGATGATGAGAGAGAGAGAGATTCAT	772
QY	1681	TCAAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGAGTTTGTGGAAGAGAGAGAGATG	1740
Db	773	GATGATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGATGATGATGATGATGATGATG	832
QY	1741	CTGATTAAGAAGTCAAGAT	1800
Db	833	AAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	892
QY	1801	TTTGATCTGGAGAA	1814
Db	893	GATGATGAGAGAA	906

RESULT 15
AAx83006/c
ID AAX83006 standard; DNA; 16442 BP

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 14:52:55 ; Search time 7073.65 Seconds
(without alignments)
11507.244 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878

Sequence: 1 atgagctcctagcgtcgtcc.....ctcacatgaagatcga 1878

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1878	100.0	1878	6	AX078761	AX078761 Sequence
2	1878	100.0	1909	8	BT004380	BT004380 Arabidops
3	1878	100.0	2162	8	BT002944	BT002944 Arabidops
4	1462	77.8	2254	8	AF239719	AF239719 Arabidops
5	1462	77.8	3275	6	AX078760	AX078760 Sequence
6	1462	77.8	81365	8	AB025633	AB025633 Arabidops
7	437.8	23.3	650	8	ATH528171	ATH528171 Arabidops
8	427.4	22.8	2181	8	AK064995	AK064995 Oryza sat
9	372	19.8	2570	8	AK100699	AK100699 Oryza sat
10	368	19.6	1556	8	AK064217	AK064217 Oryza sat
11	291.6	15.5	1564	8	AF542974	AF542974 Trilicium
12	288.8	15.4	1214	8	AK061975	AK061975 Oryza sat
13	192	10.2	513	8	AF469493	AF469493 Trilicium
14	169.4	9.0	91053	8	CNS08CAO	AL831811 Oryza sat
15	169.4	9.0	11231	8	CNS07YQ1	AL713901 Oryza sat
16	130.4	6.9	94911	8	ATT14E10	AL138656 Arabidops
17	88.4	4.7	179685	10	AC122056	AC122056 Mus muscu
18	87.8	4.7	7218	6	I66494	I66494 Sequence 14
19	86.6	4.6	204635	10	AL928594	AL928594 Mouse DNA
20	84	4.5	282861	2	AC111701	AC111701 Rattus no
21	83.4	4.4	227272	2	AC103478	AC103478 Rattus no
22	83.2	4.4	279587	2	AC112408	AC112408 Rattus no
23	82.4	4.3	157393	10	AC124199	AC124199 Mus muscu
24	80.2	4.3	248672	2	AC105804	AC105804 Rattus no
25	79.4	4.2	124244	10	AL929026	AL929026 Mouse DNA
26	79.4	4.2	188782	10	AL713870	AL713870 Mouse DNA
27	79	4.2	195690	10	AC098726	AC098726 Mus muscu
28	78.6	4.2	114848	9	AC018664	AC018664 Homo sapi
29	78.6	4.2	178635	9	AC008080	AC008080 Homo sapi
30	78.6	4.2	194006	10	AL732441	AL732441 Mouse DNA
31	78	4.2	233713	2	AC133034	AC133034 Rattus no
32	78	4.2	237549	2	AC112451	AC112451 Rattus no
33	77.8	4.1	201074	2	AC102097	AC102097 Mus muscu
34	77.8	4.1	211947	10	AC107828	AC107828 Mus muscu
35	77.6	4.1	188587	10	AC124464	AC124464 Mus muscu
36	77.6	4.1	205037	10	AC126266	AC126266 Mus muscu
37	77.4	4.1	169256	10	AC127264	AC127264 Mus muscu
38	77.4	4.1	178927	2	AC118897	AC118897 Rattus no
39	77.4	4.1	217511	10	AC125357	AC125357 Rattus no
40	77.2	4.1	56052	8	AP006363	AP006363 Lotus cor
41	77	4.1	215910	10	AC127335	AC127335 Mus muscu
42	76.8	4.1	195987	10	AC122537	AC122537 Mus muscu
43	76.8	4.1	197599	2	AC119769	AC119769 Rattus no
44	76.8	4.1	226601	10	AC122863	AC122863 Mus muscu
45	76.6	4.1	248718	10	AC117629	AC117629 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS AX078761 1878 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 2 from Patent WO01005951.
ACCESSION AX078761
VERSION AX078761.1 GI:13158380
SOURCE
KEYWORDS
ORGANISM Arabidopsis thaliana (chale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS Beclin,C., Elmayan,T. and Vaucheret,H.
TITLE Novel sgsl plant gene and use thereof

JOURNAL

Patent: WO 0105951-A 2 25-JAN-2001;
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
 AGRONOMIQUE (FR)

FEATURES

Source

Location/Qualifiers

1..1878

/organism="Arabidopsis thaliana"

/mol_type="unassigned DNA"

/db_xref="taxon:3702"

CDS

1..1878

/note="unnamed protein product"

/codon_start=1

/protein_id="CAC32419.1"

/db_xref="GI:13158381"

/db_xref="EMBL: CAC32419"

translation="MSSRAQPMSEKENTVGGYRPEVQLVQGLAGTRLIASSQDDGEM
 EYISKNNKPKPNTSGKTWVSONSNPRAVSGQOQSGNSVNGNNVSGNGNGRG
 IOANTSGRGALSRVYDNNFVA PVPSRPLEGNNWMOARGSAOHTVQFPEVED
 VNASEBENDSDALDDSDDDLASDSDSDVSKSHSGKONKFKFSGLSLSEIG
 INEPQRMWHPACQNGPGALIDMTNLHPILAHARTGARVYLHPRLAEKDLIMRG
 ASVIPCGLYQWKGIGDEKDYELVWPPVITINTRLDKDNDKMLQMGQELLEYP
 DYEALRAHSHYPOGHGMSVLMFESSATGYLAEHLRLHRLAEGLDRLIAMQKQSM
 PGGVROLYGLATKODDIDFNOHSGGTRLFELKSYQEMVWKELRQISDNOQNTY
 PNKXLSKONKHAKVLBSLEIMSEKLRTAEEDNRIVROTKMHEKONBEADHRFP
 MSLKQIHRBRPAKEENPEMLQOQERAVKVGQOQONINPSSNDCKRKAEEVSEIF
 QKEMEFEVEREMILKQEKQEMDKRKHHEIFDLKEPDEALQOLMYKHGLNED
 D"

ORIGIN

Query Match 100.0%; Score 1878; DB 6; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTTCTAGAGGCTGTCATATGTCTAAGAAAAGACGTTGAGGTTATAGGCT 60
 1 ATGAGTTCTAGAGGCTGTCATATGTCTAAGAAAAGACGTTGAGGTTATAGGCT 60
 61 GAGGTTGAACAGTTGGTTCAAGGTTGGCAGGAGCAGACGTCCTTCAACAAGATAT 120
 61 GAGGTTGAACAGTTGGTTCAAGGTTGGCAGGAGCAGACGTCCTTCAACAAGATAT 120
 121 GAGAGAGAGTGGAGGTCATTTCCAGAGAGACAAAGAACAAACAGAGAACTTTGGA 180
 121 GAGAGAGAGTGGAGGTCATTTCCAGAGAGACAAAGAACAAACAGAGAACTTTGGA 180
 181 AAAAATTGGGTTCTCAAAATTCCTCTAGAGCTTGGGTTGTCAGCAGCAAGG 240
 181 AAAAATTGGGTTCTCAAAATTCCTCTAGAGCTTGGGTTGTCAGCAGCAAGG 240
 241 AGAGGTGCAACGATCTGGAGAGAGAAACATGTATCCGGAGAGGTTAACGGCAATGT 300
 241 AGAGGTGCAACGATCTGGAGAGAGAAACATGTATCCGGAGAGGTTAACGGCAATGT 300
 301 CGGGGCAATTCAGGTTCAATATCTGATCGGGGAGCAGCGTTGAGCAGAAAGTATGATAC 360
 301 CGGGGCAATTCAGGTTCAATATCTGATCGGGGAGCAGCGTTGAGCAGAAAGTATGATAC 360
 361 AACTTTGTGGCAACCCCACTGTATCTGCGCTCTTGGAAAGAGAGATGGAATTTGGCAG 420
 361 AACTTTGTGGCAACCCCACTGTATCTGCGCTCTTGGAAAGAGAGATGGAATTTGGCAG 420
 421 GCAAGAGAGAGGTTCTGCTCAGCAGACAGCTGTGAGAGAGTTTCTGACGTGAGAGATAT 480
 421 GCAAGAGAGAGGTTCTGCTCAGCAGACAGCTGTGAGAGAGTTTCTGACGTGAGAGATAT 480
 481 GTGATAATGCTTCTGAGAGAGAGATATCCGATGCTTTGATGATATTCGATGAGAC 540
 481 GTGATAATGCTTCTGAGAGAGAGATATCCGATGCTTTGATGATATTCGATGAGAC 540
 541 CTTGCAAGATGATATGATGATCGGATGAGTGAAGAGCAATGATGATGATGATGATGATGAT 600
 541 CTTGCAAGATGATATGATGATCGGATGAGTGAAGAGCAATGATGATGATGATGATGATGAT 600
 601 AATAAGTGTTCAAAAGTTCTTTGGCAGCTTGATGATGATGATGATGATGATGATGATGAT 660

601 AATAAGTGTTCAAAAGTTCTTTGGCAGCTTGATGATGATGATGATGATGATGATGATGAT 660
 661 GAACCAAGAGGAGATGGCATTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 661 GAACCAAGAGGAGATGGCATTTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 720
 721 TATACCTGAGACCTCTACAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 721 TATACCTGAGACCTCTACAGCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 780
 781 CATAGAAATTTGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAG 840
 781 CATAGAAATTTGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAG 840
 841 CTTGTGTGATGATTTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 841 CTTGTGTGATGATTTATGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
 901 ATTGCTGGGCTCCATGATGATCATCATGATGATGATGATGATGATGATGATGATGAT 960
 901 ATTGCTGGGCTCCATGATGATCATCATGATGATGATGATGATGATGATGATGATGAT 960
 961 AAGTGGCTGGCATGGGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 961 AAGTGGCTGGCATGGGCAACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
 1021 AGAGAGGCGCATTCCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 1021 AGAGAGGCGCATTCCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
 1081 AGCAGTGCACATGCTGATTTGAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 1081 AGCAGTGCACATGCTGATTTGAGAGGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
 1141 TTAGTAAATTTGCTGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 1141 TTAGTAAATTTGCTGGGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
 1201 TATGCTCTCTTGCACAG 1260
 1201 TATGCTCTCTTGCACAG 1260
 1261 ACAAGGCTGAATTTGAG 1320
 1261 ACAAGGCTGAATTTGAG 1320
 1321 ATCTCTGAG 1380
 1321 ATCTCTGAG 1380
 1381 CAGGCAAGAGTCTTGAAGAGATCTCTGAAATTTAGAGAGAGAGAGAGAGAGAGAG 1440
 1381 CAGGCAAGAGTCTTGAAGAGATCTCTGAAATTTAGAGAGAGAGAGAGAGAGAGAG 1440
 1441 GAGATATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 1441 GAGATATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
 1501 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 1501 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
 1561 GCAAGAGAGAGAGATTTGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 1561 GCAAGAGAGAGAGATTTGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 1621 CAGCAGAGAGAGATTTGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
 1621 CAGCAGAGAGAGATTTGAGAGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
 1681 TCAAGCTTCATGAGATTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740

Db 1681 TCAAGCTTCATGAGCTTTCAGAGAAAGATGAGAGAGCTTTGTGAGAGAGAGGAGATG 1740

QY 1741 CTGATTAAGATCAAGAGAGATGAGAGATGAGATGAGATGAGATGAGATGAGATG 1800

Db 1741 CTGATTAAGATCAAGAGAGATGAGAGATGAGATGAGATGAGATGAGATGAGATG 1800

QY 1801 TTTGATCTGAGAGAAAGATTTGATGAGCTTTGAGAACAGCTGATGACAGATGAGCTT 1860

Db 1801 TTTGATCTGAGAGAAAGATTTGATGAGCTTTGAGAACAGCTGATGACAGATGAGCTT 1860

QY 1861 CACATGAGAGATGATGGA 1878

Db 1861 CACATGAGAGATGATGGA 1878

RESULT 2

LOCUS BT004380 1909 bp mRNA linear PLN 14-FEB-2003

DEFINITION Arabidopsis thaliana clone U20243 unknown protein (At5g23570) mRNA, complete cds.

ACCESSION BT004380

VERSION BT004380.1 GI:28393932

KEYWORDS FLI CDNA.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsids.

1 (bases 1 to 1909)

REFERENCE

AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Tortum,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Arabidopsis Open Reading Frame (ORF) Clones

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1909)

AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Tortum,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

TITLE Direct Submission

JOURNAL Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA

COMMENT The RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai, J., Hayashizaki,Y. and Shinozaki,K.

The Sak, Stanford, PGEC (SSP) Consortium members constructed and sequenced the pUNI (ORF) clones using the RAFL cDNAs: Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Tortum,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.

FEATURES

source location/Qualifiers

1..1909

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

gene

CDS

1..1909

/gene="At5g23570"

1..1878

/gene="At5g23570"

/codon_start=1

/evidence=experimental

/product="unknown protein"

/protein_id="AA042374.1"

/db_xref="GI:28393933"

/translation="MSSRAGPMSKENVYGGYRPEVVDLVAGTRLASDDDGEW EVISKKNKRGNTSGKTWVGNSSNPPRAMPKQOQGRGNSVGRNNVGRNGRG IOANTSGRRLSRKYDNNFVAPPPVSPPLEGGNNQKRGSAOHTVAOEPDVEDD VDNASEENDSDALDDSDLDASDDYSDVSQKSHSRKQNKPKFKPSLDSLS1EQ INEPORWMCAPQKNGPAGIDWYNIPLAHARTGABRVKHLRELAELERDLOMRG ASVLPCEIYGOMKGLGDEDEKYEIVMPPMVYINMTRLDKNDKMLGNGQLELYF DKVEALRARHSYGPQHRGMSVLMPESSATGYLAEARLAEVEMVRELROISNOLNY FGNLSKQNKAKVLEBSLEIMSEKLRRTAEDNRIVRQRTKQHEQNBEMADHRRF MDSIKQIHERDARENEEMLOQQRARVVGQOQNIINSSNDCKRKAEEVSFTIE OEKEMEFEVEEREMLIKOEKMEKEMKRRHHEIFDLKEFDEALEQLMYKGLHNE D"

1879..1909

/gene="At5g23570"

ORIGIN

3'UTR

Query Match

Best Local Similarity 100.0%; Score 1878; DB 8; Length 1909;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAGTTCTAGGCGCTGCTCAATGCTTAAAGAAAGAACGTTCAAGGCTTATAGCCT 60

1 ATGAGTTCTAGGCGCTGCTCAATGCTTAAAGAAAGAACGTTCAAGGCTTATAGCCT 60

QY 61 GAGGTTGAACGTTGCTTCAAGCTTTGGCAGGAGAGAGCTGCTTCAACAATGAT 120

Db 61 GAGGTTGAACGTTGCTTCAAGCTTTGGCAGGAGAGAGCTGCTTCAACAATGAT 120

QY 61 GAGGTTGAACGTTGCTTCAAGCTTTGGCAGGAGAGAGCTGCTTCAACAATGAT 120

Db 61 GAGGTTGAACGTTGCTTCAAGCTTTGGCAGGAGAGAGCTGCTTCAACAATGAT 120

QY 121 GGAGGAGAGTGGAGAGTCAATTTCCAAAGAAACAACAACGAGAAACCTTCTGGA 180

Db 121 GGAGGAGAGTGGAGAGTCAATTTCCAAAGAAACAACAACGAGAAACCTTCTGGA 180

QY 121 GGAGGAGAGTGGAGAGTCAATTTCCAAAGAAACAACAACGAGAAACCTTCTGGA 180

Db 121 GGAGGAGAGTGGAGAGTCAATTTCCAAAGAAACAACAACGAGAAACCTTCTGGA 180

QY 181 AAAAAGGCTTCTCAAGATTCGATCTTCTTCAAGCTTGGGAGTGGTCAAGAGAGG 240

Db 181 AAAAAGGCTTCTCAAGATTCGATCTTCTTCAAGCTTGGGAGTGGTCAAGAGAGG 240

QY 181 AAAAAGGCTTCTCAAGATTCGATCTTCTTCAAGCTTGGGAGTGGTCAAGAGAGG 240

Db 181 AAAAAGGCTTCTCAAGATTCGATCTTCTTCAAGCTTGGGAGTGGTCAAGAGAGG 240

QY 241 AGAGTTACCAAGTATCTGGGAGAGAAACAATGATTCGGAGAGGATGATGATGAT 300

Db 241 AGAGTTACCAAGTATCTGGGAGAGAAACAATGATTCGGAGAGGATGATGATGAT 300

QY 241 AGAGTTACCAAGTATCTGGGAGAGAAACAATGATTCGGAGAGGATGATGATGAT 300

Db 241 AGAGTTACCAAGTATCTGGGAGAGAAACAATGATTCGGAGAGGATGATGATGAT 300

QY 301 CGGGGCTTCAAGCTTCAATATCTGCTGGGAGAGAGGCTTGAAGCAAAAGTATGAT 360

Db 301 CGGGGCTTCAAGCTTCAATATCTGCTGGGAGAGAGGCTTGAAGCAAAAGTATGAT 360

QY 301 CGGGGCTTCAAGCTTCAATATCTGCTGGGAGAGAGGCTTGAAGCAAAAGTATGAT 360

Db 301 CGGGGCTTCAAGCTTCAATATCTGCTGGGAGAGAGGCTTGAAGCAAAAGTATGAT 360

QY 361 AACTTTGTGGACACCCCACTGTATCTGCGCTCTTTGGAAGAGAGATGATGGCAG 420

Db 361 AACTTTGTGGACACCCCACTGTATCTGCGCTCTTTGGAAGAGAGATGATGGCAG 420

QY 361 AACTTTGTGGACACCCCACTGTATCTGCGCTCTTTGGAAGAGAGATGATGGCAG 420

Db 361 AACTTTGTGGACACCCCACTGTATCTGCGCTCTTTGGAAGAGAGATGATGGCAG 420

QY 421 GCAAGAGAGGTTCTGCTCAGACACACCTGTGAGAGAGTTCCTGAGCTGAGATGAT 480

Db 421 GCAAGAGAGGTTCTGCTCAGACACACCTGTGAGAGAGTTCCTGAGCTGAGATGAT 480

QY 421 GCAAGAGAGGTTCTGCTCAGACACACCTGTGAGAGAGTTCCTGAGCTGAGATGAT 480

Db 421 GCAAGAGAGGTTCTGCTCAGACACACCTGTGAGAGAGTTCCTGAGCTGAGATGAT 480

QY 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGATGATGATGATGATGATGAT 540

Db 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGATGATGATGATGATGATGAT 540

QY 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGATGATGATGATGATGATGAT 540

Db 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGATGATGATGATGATGATGAT 540

QY 541 CTTCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

Db 541 CTTCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600

QY	601	ATAAGTGGTCAAAAAGTCTTTGGCACTTGGATAGCTTGTGCATCGACGATTAAT	660
Db	601	ATAAGTGGTCAAAAAGTCTTTGGCACTTGGATAGCTTGTGCATCGACGATTAAT	660
QY	661	GAACACAGAGGAGTGGCATTTGTCAGCTGTGCAGAACGACCTGGTGCATCGATTGG	720
QY	661	GAACACAGAGGAGTGGCATTTGTCAGCTGTGCAGAACGACCTGGTGCATCGATTGG	720
QY	721	TATTAACCTGCACCTCTACTAGCTCATCGAGAGCAAAAAGAGCTAGACGATTAGCTC	780
Db	721	TATTAACCTGCACCTCTACTAGCTCATCGAGAGCAAAAAGAGCTAGACGATTAGCTC	780
QY	781	CATAGAGAAATGGCTGAAATTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGCAAT	840
Db	781	CATAGAGAAATGGCTGAAATTTTAAAGAAAGATCTACAGATGAGAGCGCATCTGCAAT	840
QY	841	CCTTGTGTGAGATTTATGAGGAGTGGAGGCTTGGGAGATGATAAGATTTATGAA	900
Db	841	CCTTGTGTGAGATTTATGAGGAGTGGAGGCTTGGGAGATGATAAGATTTATGAA	900
QY	901	ATTGTCTGGCTCCAAATGATCATCATGAAATCTAAGCTGATAGGACATTAACGAT	960
Db	901	ATTGTCTGGCTCCAAATGATCATCATGAAATCTAAGCTGATAGGACATTAACGAT	960
QY	961	AAGTGGCTCGGCATGGGCAACCAAGAGCTGTGGAATTACTTGAACAAGATGAGGCTCTT	1020
Db	961	AAGTGGCTCGGCATGGGCAACCAAGAGCTGTGGAATTACTTGAACAAGATGAGGCTCTT	1020
QY	1021	AGAGAGCCCATTCCTTAATGATCCACAGAGGCAATGAGGATGATGATTTGATGTTGAG	1080
Db	1021	AGAGAGCCCATTCCTTAATGATCCACAGAGGCAATGAGGATGATGATTTGATGTTGAG	1080
QY	1081	AGCAGTGCACATGCTATTTGAGAGCCCAAGAGCTCCACCGGAGTTAGCTGAGATGGG	1140
Db	1081	AGCAGTGCACATGCTATTTGAGAGCCCAAGAGCTCCACCGGAGTTAGCTGAGATGGG	1140
QY	1141	TTAGATGAATTTGCTGGGGGTGAGAGGCGAGTATGTTTGTGAGGTTTCCGCAATG	1200
Db	1141	TTAGATGAATTTGCTGGGGGTGAGAGGCGAGTATGTTTGTGAGGTTTCCGCAATG	1200
QY	1201	TATGCTTCTCTGCAACGAGCAAGCAAGTCTGCAATTAACAACCTCTCAAGGCAAA	1260
Db	1201	TATGCTTCTCTGCAACGAGCAAGCAAGTCTGCAATTAACAACCTCTCAAGGCAAA	1260
QY	1261	ACAAGCTGAATTCGATGGAATCATACCAAGAGTGTGTAAGAGAGCTGAGGAG	1320
Db	1261	ACAAGCTGAATTCGATGGAATCATACCAAGAGTGTGTAAGAGAGCTGAGGAG	1320
QY	1321	ATCTCTGAGGACATTCAGAGCTGAACTCTTTAAGAACAGCTCTCAAAACAGAACAG	1380
Db	1321	ATCTCTGAGGACATTCAGAGCTGAACTCTTTAAGAACAGCTCTCAAAACAGAACAG	1380
QY	1381	CACGCCAAGGCTGTGAGGATCTCTGGAATTAATAGCGAGAGAGCTGGTGAAGTCTGA	1440
Db	1381	CACGCCAAGGCTGTGAGGATCTCTGGAATTAATAGCGAGAGAGCTGGTGAAGTCTGA	1440
QY	1441	GAGGATTAATCGATGAGAGCAAGAACTTAAGATGAGATGAACAGAACAGAGAGAG	1500
Db	1441	GAGGATTAATCGATGAGAGCAAGAACTTAAGATGAGATGAACAGAACAGAGAGAG	1500
QY	1501	ATGATGACACAGAGGTTTTCATGATTCATCAACAGATTCATGAAGAGAGAGC	1560
Db	1501	ATGATGACACAGAGGTTTTCATGATTCATCAACAGATTCATGAAGAGAGAGC	1560
QY	1561	GCAGAGAGAGAAATTCAGATGTTGACAGAGCAAGAGCTGCAAGAGTTGTCAGAG	1620
Db	1561	GCAGAGAGAGAAATTCAGATGTTGACAGAGCAAGAGCTGCAAGAGTTGTCAGAG	1620
QY	1621	CAGCAGAGAGAAATTCAGATGTTGACAGAGCAAGAGCTGCAAGAGTTGTCAGAG	1680
Db	1621	CAGCAGAGAGAAATTCAGATGTTGACAGAGCAAGAGCTGCAAGAGTTGTCAGAG	1680

QY	1681	TCAGCTTCATGAGTTTCAAGAGAAAGATGAGAGGAGTTTGTGAGAGAGAGGAGATG	1740
Db	1681	TCAGCTTCATGAGTTTCAAGAGAAAGATGAGAGGAGTTTGTGAGAGAGAGGAGATG	1740
QY	1741	CTGATTAAGATCAGAGAGAGAGATGAGAGCATGAGAGAGGATCAGAGAGATTA	1800
Db	1741	CTGATTAAGATCAGAGAGAGAGATGAGAGCATGAGAGAGGATCAGAGAGATTA	1800
QY	1801	TTGATCTGAGAGAAAGATTTGATGAGGCTTTGGAACAGCTCATATGACAAGCATG	1860
Db	1801	TTGATCTGAGAGAAAGATTTGATGAGGCTTTGGAACAGCTCATATGACAAGCATG	1860
QY	1861	CACATGAGAGATGATGA	1878
Db	1861	CACATGAGAGATGATGA	1878

RESULT 3	BT002944	2162 bp	mRNA	linear	PLN 15-JAN-2003
LOCUS	BT002944				
DEFINITION	Arabidopsis thaliana clone RAFL4-93-K05 (R20243) unknown protein				
ACCESSION	BT002944	GI:27754622			
VERSION	BT002944.1				
KEYWORDS	FLI CDNA.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carrinchi, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.				
AUTHORS	Arabidopsis Full Length cDNA Clones				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2162)				
AUTHORS	Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carrinchi, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA (GSC) members carried out the				
COMMENT	RIKEN Genomic Sciences Center collection and clustering of RAFL cDNAs (RAFL CDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki, M., Narusaka, M., Ishida, J., Satou, M., Kamiya, A., Sakurai, T., Carrinchi, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K.				

The Salik, Stanford, PEGC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs. Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G., Yuan, S., Carrinchi, P., Chen, H., Cheuk, R., Hayashizaki, Y., Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M., Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PEGC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PEGC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.

Location/Qualifiers

1..2162


```

/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAF14-93-K05 (R20243)"
/ecotype="Columbia"
/notes="This clone is in a modified pBluescript vector2
(lambda PS) as a BamHI/XhoI insert."
1..2162
/gene="At5g23570"
1..91
/gene="At5g23570"
92..1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AA022757.1"
/db_xref="GI:27754623"
/translation="MSSRAQPMKSEKNYVGYPREVDLVGLAGTRLASDQDQGM
EVIKKNKPKGNTSGKTWVSONSNPPRPMGQOQGRSNTSGRNNVGRNGRG
IOANISGRBALSRKYDNNFYAPPVSRPLEGGMNQARGSAAHTVAQEPDVEDD
VNASSEBNDSDALDDDDLASDDYDSVSKSHSRKQNKWFKKFVGLDLSIEQ
INBPQRMHCPAQCPQPGAIQWYVNLPLAHARTGARVXLHRELAVLEKDLQMRG
ASVIPCGBIYQGMKGLGDEKDYELVMPMVYINMTRLDKQNDKMLGKQDLLEYF
DKYEALRAHSYGPQGHRSYVLMFESSATGLAEARLHRELAEGLDRIAMGQRSM
PSGVRQLYGLATKODLDIPNOSHOGTRLKEIKSYQEWVYKELROIISDNOLNY
FKNTLSKQNKHAKVLEBSLEINSEKLRTRAEENRTVROPTKQHQONBENDHRRPF
MDSIKQIHERNDKSENFEMLOQDERAVGQOQONIPSSNDCKRAEVSSTIEF
QKEMEEFVEEREMLIKQEKEMEDKRRHHEIFDLKEFDEALEOLMYKRLHNEED
D"
1970..2162
/gene="At5g23570"
2146
/gene="At5g23570"
/notes="not present in genomic sequence"

ORIGIN
Query Match 100.0%; Score 1878; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTTAGGGCTGTGTCATGTCTAAGAAAAGAAAGTTCAAGGTTGTTATAGGCT 60
DB 92 ATGAGTTTAGGGCTGTGTCATGTCTAAGAAAAGAAAGTTCAAGGTTGTTATAGGCT 151
QY 61 GAGGTTGAACAGTTGTTCAAGGTTTGCGAGGACGAGACTGGCTTTTCAACAATGAT 120
DB 152 GAGGTTGAACAGTTGTTCAAGGTTTGCGAGGACGAGACTGGCTTTTCAACAATGAT 211
QY 121 GAGGAGAGTGGAGAGTCAATTTCCAAAGAAACAAGAAACAACCGAGAAACCTCTGGA 180
DB 212 GAGGAGAGTGGAGAGTCAATTTCCAAAGAAACAAGAAACAACCGAGAAACCTCTGGA 271
QY 181 AAAAATTGGGTTTCTCAGAAATTCGAATCCTCTAGAGCTTGGGTTGTCAGACGAAAGG 240
DB 272 AAAAATTGGGTTTCTCAGAAATTCGAATCCTCTAGAGCTTGGGTTGTCAGACGAAAGG 331
QY 241 AGAGGTAGCAACGTAATCTGGGAGAGAAACAATGTATCCGGGAGAGTTAACCGCAATGAT 300
DB 332 AGAGGTAGCAACGTAATCTGGGAGAGAAACAATGTATCCGGGAGAGTTAACCGCAATGAT 391
QY 301 CGGGGCAATTCAGACGTAATCTGGTGGGAGACGAGCTTGAAGCAAAAGTATGATTAAC 360
DB 392 CGGGGCAATTCAGACGTAATCTGGTGGGAGACGAGCTTGAAGCAAAAGTATGATTAAC 451
QY 361 AACTTTGTGGACCCCACTGTATCTGCCCTCTCTTTGGAAGAGAGATGAATTTGGCAG 420
DB 452 AACTTTGTGGACCCCACTGTATCTGCCCTCTCTTTGGAAGAGAGATGAATTTGGCAG 511
QY 421 GCAAGAGAGAGTTGCTCAGACACAGCTGTGCGAGAGTTTCTTGACGTGAGGATGAT 480
DB 512 GCAAGAGAGAGTTGCTCAGACACAGCTGTGCGAGAGTTTCTTGACGTGAGGATGAT 571

```

```

QY 481 GTGATTAATGCTTCTGAGGAAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 540
DB 572 GTGATTAATGCTTCTGAGGAAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 631
QY 541 CTGGCAAGTATGATTAATGATCTGGATGTGACTCAAAAGAGCCATGATCAAGAAAGCAG 600
DB 632 CTGGCAAGTATGATTAATGATCTGGATGTGACTCAAAAGAGCCATGATCAAGAAAGCAG 691
QY 601 AATAAGTGTCAAAAGTCTTTGGCAGCTTGGATGATCTGTGCATGAGAGATTAAT 660
DB 692 AATAAGTGTCAAAAGTCTTTGGCAGCTTGGATGATCTGTGCATGAGAGATTAAT 751
QY 661 GAACCAAGAGGACAGTGCATTTGTCACGTTGTCAAGAACGACCTGTGCCATTTGG 720
DB 752 GAACCAAGAGGACAGTGCATTTGTCACGTTGTCAAGAACGACCTGTGCCATTTGG 811
QY 721 TATAAAGTGCACCTCTCTAGCTGATGCGAGACAAAGAGCTAAGCGATTAAGCTC 780
DB 812 TATAAAGTGCACCTCTCTAGCTGATGCGAGACAAAGAGCTAAGCGATTAAGCTC 871
QY 781 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGTCAAT 840
DB 872 CATAGAGATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGGCGCATCTGTCAAT 931
QY 841 CTTGTGGGAGATTTATGGGCAGTGAAGGGTTTGGGAGAGATGAAGAAAGATTAATGA 900
DB 932 CTTGTGGGAGATTTATGGGCAGTGAAGGGTTTGGGAGAGATGAAGAAAGATTAATGA 991
QY 901 ATTGTCTGGCTCCAAATGTCATCATCATGATTAAGATTAAGACGATTAACGAT 960
DB 992 ATTGTCTGGCTCCAAATGTCATCATCATGATTAAGATTAAGACGATTAACGAT 1051
QY 961 AAGTGGCTTCGGCATGAGGCAACCAAGAGCTGTGGAATCTTGCACAAAGTATGAGCTCTT 1020
DB 1052 AAGTGGCTTCGGCATGAGGCAACCAAGAGCTGTGGAATCTTGCACAAAGTATGAGCTCTT 1111
QY 1021 AGAGCAAGCCAAATCTCTATGTCACACAGGAGCCATCGTGGAGATGAGTCTGATTTTGA 1080
DB 1112 AGAGCAAGCCAAATCTCTATGTCACACAGGAGCCATCGTGGAGATGAGTCTGATTTTGA 1171
QY 1081 AGCAGTGCACATGCTATTTTGAAGGCCGAAACGCTTCCACCGGAGTTAGCTGATGAGG 1140
DB 1172 AGCAGTGCACATGCTATTTTGAAGGCCGAAACGCTTCCACCGGAGTTAGCTGATGAGG 1231
QY 1141 TTAGATTAATTTGCTGGGCTCAAGAGCGCAGTATGTTTTCTGAGAGTTCGCCAATCG 1200
DB 1232 TTAGATTAATTTGCTGGGCTCAAGAGCGCAGTATGTTTTCTGAGAGTTCGCCAATCG 1291
QY 1201 TATGCTCTCTTGCAGAGAGAGATCTGGACATTAATCAATCAACACTCTCAAGGCAAA 1260
DB 1292 TATGCTCTCTTGCAGAGAGAGATCTGGACATTAATCAATCAACACTCTCAAGGCAAA 1351
QY 1261 ACAAGCTGAATTCGATGTAATCATACCAAGAGATGTTGTAAGAGCTGAGGCGAG 1320
DB 1352 ACAAGCTGAATTCGATGTAATCATACCAAGAGATGTTGTAAGAGCTGAGGCGAG 1411
QY 1321 ATCTTGAAGCAATCAGACGTGAATCTTTAAAGAACAGCTTCAAAACAGAACAG 1380
DB 1412 ATCTTGAAGCAATCAGACGTGAATCTTTAAAGAACAGCTTCAAAACAGAACAG 1471
QY 1381 CACGCAAGGTGCTTGAAGAAATCTCTGGAATTTAGACGAGAAAGCTGCTGTAAGCTGA 1440
DB 1472 CACGCAAGGTGCTTGAAGAAATCTCTGGAATTTAGACGAGAAAGCTGCTGTAAGCTGA 1531
QY 1441 GAGGATTAATCGGATCTGTGAGACAGAGAACTAAGATGACAGATGAACAGAGGAAAG 1500
DB 1532 GAGGATTAATCGGATCTGTGAGACAGAGAACTAAGATGACAGATGAACAGAGGAAAG 1591
QY 1501 ATGATGCAACACAGAGTCTTTTCAATGATTAATCAAAACAGATTCATGAAGAAAGAGAC 1560
DB 1592 ATGATGCAACACAGAGTCTTTTCAATGATTAATCAAAACAGATTCATGAAGAAAGAGAC 1651

```

OY	1561	GCAAAAGGAGAATTTTCGAGATGTTCACGCACACAGAACCTGCCAAGTGTGTGGCCAG	1620
Db	1652	GCAAAAGGAGAATTTTCGAGATGTTTGACACACAGAAAGTGTCCAAGTGTGTGGCCAG	1711
OY	1621	CAGCAGCAGAACATTAAATCCCTCTAGGAATGACGATGCGCAAAGAGACTGAGAAAGTG	1680
Db	1712	CAGCAGCAGAACATTAAATCCCTCTAGCAATGACATTTGCCGAAAAGAGACTGAGAAAGTG	1771
OY	1661	TCAAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGATTGTGTGAAAGAGAGGAGATG	1740
Db	1772	TCACGCTTCATCGAGTTTCAAGAGAAAGATGAGAGATTGTGTGAAAGAGAGGAGATG	1831
OY	1741	CTGATTAAGATCAAGAGAAAGATGGAACAATGAAGAGAGCATCACAGAGAGATA	1800
Db	1832	CTGATTAAGATCAAGAGAAAGATGGAAGACATGAAAGAGAGCATCACAGAGAGATA	1891
OY	1801	TTTGATCTGGAGAAGAAATTTGATGAGGCTTTGAAACAGCTCATGTATCAACACATGGGCTT	1860
Db	1892	TTTGATCTGGAGAAGAAATTTGATGAGGCTTTGAAACAGCTCATGTATCAACACATGGGCTT	1951
OY	1861	CACAAATGAAGATGATTGA	1878
Db	1952	CACAAATGAAGATGATTGA	1969

RESULT 4	AF239719	2254 bp	DNA	linear	PLN 02-JUN-2001
LOCUS	AF239719				
DEFINITION	Arabidopsis thaliana SGS3 gene, complete cds.				

QY 961 A----- 961
 Db 961 AAGTGAATCTCTGCTTTTACTTTTATTTTCTTCATCTTACTGATCTTA 1020
 QY 962 -----AGTGGCTGGCATGGGCAACCAAGCTGCTGGAATCTTGACAA 1007
 Db 1021 GAATGTATCATTTGTAGTGGCTGGGCAACCAAGCTGCTGGAATCTTGACAA 1080
 QY 1008 GTATGAGGCTTTAGAGCAGCGCATTCATGCTCCAGGGCCATCGGGATGAGTGT 1067
 Db 1081 GTATGAGGCTTTAGAGCAGCGCATTCATGCTCCAGGGCCATCGGGATGAGTGT 1140
 QY 1068 TCTGATGTTTGAAGCAGCGCATTCATGCTCCAGGGCCATCGGGATGAGTGT 1127
 Db 1141 TCTGATGTTTGAAGCAGCGCATTCATGCTCCAGGGCCATCGGGATGAGTGT 1200
 QY 1128 AGCTGAGATGGGCTTATGATGAATTCCTGGGGTCCAGAACCGCATGTTTCTGGAG 1187
 Db 1201 AGCTGAGATGGGCTTATGATGAATTCCTGGGGTCCAGAACCGCATGTTTCTGGAG 1260
 QY 1188 TGTGGCCAACTGTATGCTTCTTGCAACGAAGCATTCGACATTTCAATCAACA 1247
 Db 1261 TGTGGCCAACTGTATGCTTCTTGCAACGAAGCATTCGACATTTCAATCAACA 1320
 QY 1248 CTCTCA----- 1253
 Db 1321 CTCTCAAGTCTCTCCCAAGAAATTTGATATATGCTTTAGTTTGTCAATGGAAT 1380
 QY 1254 -----AGG 1256
 Db 1381 TTAAGTTTGTGTGCTGTTAATGATCATCTGTTATGATATATATGATCATTTAG 1440
 QY 1257 CAACCAAGGCTGAATTCGATTTGAATATCAATCAAGAGATGTTGTAAGAAGTGG 1316
 Db 1441 CAACCAAGGCTGAATTCGATTTGAATATCAATCAAGAGATGTTGTAAGAAGTGG 1500
 QY 1317 GCAGATCTCTGAGCAATCAGCAGCTGAATCTTTAAGAACAACTCTCAAAACAGA 1376
 Db 1501 GCAGATCTCTGAGCAATCAGCAGCTGAATCTTTAAGAACAACTCTCAAAACAGA 1560
 QY 1377 CAAGCAGCGCAAGTGTGAGGAATCTCTGAAATTTATGACGAGAACTCGTGAAC 1436
 Db 1561 CAAGCAGCGCAAGTGTGAGGAATCTCTGAAATTTATGACGAGAACTCGTGAAC 1620
 QY 1437 TGCAGAGATATCGGATGCTGAGCAGAGAACTAAGATGCGCATGAAACAGAGGA 1496
 Db 1621 TGCAGAGATATCGGATGCTGAGCAGAGAACTAAGATGCGCATGAAACAGAGGA 1680
 QY 1497 AG----- 1498
 Db 1681 AGAGGTATGATTTTCTAGAAAATCAAACTTGACATTTGTATTAACCTAATGATTA 1740
 QY 1499 -----AGATGATGCAACGA 1514
 Db 1741 CATTTGATATATTTGTCACAAACAAACCTGTGGTGTGGAAGATGATGACACAGA 1800
 QY 1515 CAGGTTTTTCATGATTCATCAATCAACAGATTCATGAAAGAGAGCGCAAGAGAGAA 1574
 Db 1801 CAGGTTTTTCATGATTCATCAATCAACAGATTCATGAAAGAGAGCGCAAGAGAGAA 1860
 QY 1575 TTTGAGATGTTGAGCAGCAGAGAAAGTTCGCAAGGTTGTCGAGCAGCAGAGAACT 1634
 Db 1861 TTTGAGATGTTGAGCAGCAGAGAAAGTTCGCAAGGTTGTCGAGCAGCAGAGAACT 1920
 QY 1635 TAATCCCTTAGCAATGAGATTCGCAAG----- 1665
 Db 1921 TAATCCCTTAGCAATGAGATTCGCAAGAGAGATATATGATTAATAATATTCCT 1980
 QY 1666 ----- 1665
 Db 1981 CTGGGCTTTTGTGTTTCAAACTTAAGATTAATAATATTCGAGTTTGTGATCTTTG 2040

QY 1666 -AGAGCTGAGAGAGTGTCAAGCTTCATGAGTTTCAAGAGAAAGAGATGAGAGTTGT 1724
 Db 2041 CAGAGCTGAGAGAGTGTCAAGCTTCATGAGTTTCAAGAGAAAGAGATGAGAGTTGT 2100
 QY 1725 GGAAGAGAGAGAGTGTGTATTAAGATTCAGAGAAAGAGATGAGAGATGAGAGAG 1784
 Db 2101 GGAAGAGAGAGAGTGTGTATTAAGATTCAGAGAAAGAGATGAGAGATGAGAGAG 2160
 QY 1785 GCATCAGAGAGAGATTTTATCTGAGAGAAAGATTTGATGAGGCTTTGAAACAGCTCAT 1844
 Db 2161 GCATCAGAGAGAGATTTTATCTGAGAGAAAGATTTGATGAGGCTTTGAAACAGCTCAT 2220
 QY 1845 GTACAAGCATGAGCTTCACATGAATGAATGATGA 1878
 Db 2221 GTACAAGCATGAGCTTCACATGAATGAATGATGA 2254
 RESULT 5
 AX078760 3275 bp DNA linear PAT 22-FEB-2001
 LOCUS AX078760
 DEFINITION Sequence 1 from Patent WO0105951.
 ACCESSION AX078760
 VERSION AX078760.1 GI:13158379
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1
 REFERENCE
 Beclin,C., Elmayan,T. and Vaucheret,H.
 Novel sgss plant gene and use thereof
 Parent: WO 0105951-A 1 25-JAN-2001;
 JOURNAL AVENTIS CROPS SCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
 AGRONOMIQUE (FR)
 FEATURES
 source Location/Qualifiers
 1..3275
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"
 primer_bind
 693..715
 /note="p356AD"
 primer_bind complement (2926..2952)
 primer_bind
 /note="p356Y"
 ORIGIN
 Query Match 77.8%; Score 1462; DB 6; Length 3275;
 Best Local Similarity 83.3%; Pred. No. 0;
 Matches 1878; Conservative 0; Mismatches 0; Indels 376; Gaps 4;
 QY 1 ATGAGTTCTAGAGCTGTGTCCTAATGCTTAAGAGAAAGAACTTCAGAGTGTATAGGCT 60
 Db 696 ATGAGTTCTAGAGCTGTGTCCTAATGCTTAAGAGAAAGAACTTCAGAGTGTATAGGCT 755
 QY 61 GAGTTGAAACAGTTGTTCAAGTTTGCAGAGGAGAGAGTGGCTTTCTTCACAATGAT 120
 Db 756 GAGTTGAAACAGTTGTTCAAGTTTGCAGAGGAGAGAGTGGCTTTCTTCACAATGAT 815
 QY 121 GGAGGAGAGTGGAGGTCATTTCAAGAGAAACAAGAAACAACAGAGAAACCTCTGGA 180
 Db 816 GGAGGAGAGTGGAGGTCATTTCAAGAGAAACAAGAAACAACAGAGAAACCTCTGGA 875
 QY 181 AAAAATTGGGTTTCTCAAAATTCGAATCTCTTAAGCTTGGGGTGTCAAGAGAGG 240
 Db 876 AAAAATTGGGTTTCTCAAAATTCGAATCTCTTAAGCTTGGGGTGTCAAGAGAGG 935
 QY 241 AGAGGTAGCAACGTATCTGGGAGAGAGAAACAATGATCCGGGAGAGGTAACGGCAATG 300
 Db 936 AGAGGTAGCAACGTATCTGGGAGAGAGAAACAATGATCCGGGAGAGGTAACGGCAATG 995
 QY 301 CGGGCATTTCAAGCTTAACATATCTGTGCGGAGCAGCGTTGAGCAGAAAGTATGATTAAC 360
 Db 996 CGGGCATTTCAAGCTTAACATATCTGTGCGGAGCAGCGTTGAGCAGAAAGTATGATTAAC 1055

QY 361 AACTTTGGGCAACCCCACTGTATCTCGCCCTCTTGGAGAGAGATGAAATTGGCAG 420
 Db 1056 AACTTTGGGCAACCCCACTGTATCTCGCCCTCTTGGAGAGAGATGAAATTGGCAG 1115
 QY 421 GCAAGAGAGAGTTGCTCAGCACAACAGCTGTGAGAGATTTCTGCACTGTGAGAGAT 480
 Db 1116 GCAAGAGAGAGTTGCTCAGCACAACAGCTGTGAGAGATTTCTGCACTGTGAGAGAT 1175
 QY 481 GTGATTAATGCTTTCAGAGAGAGATATCCAGTGTGAGATTTGATGATGATGAGAC 540
 Db 1176 GTGATTAATGCTTTCAGAGAGAGATATCCAGTGTGAGATTTGATGATGATGAGAC 1235
 QY 541 CTGGCAAGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1295
 Db 1236 CTGGCAAGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1295
 QY 601 AATAAGTGTTCAAAAGTTCTTTGGCAGCTTGGATAGCTTGTGATGAGAGATTAAT 660
 Db 1296 AATAAGTGTTCAAAAGTTCTTTGGCAGCTTGGATAGCTTGTGATGAGAGATTAAT 1355
 QY 661 GAACCAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 Db 1356 GAACCAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1415
 QY 721 TATAACCTGCAACCTCTACTAGCTCATGCGAGAGCAAAAGAGCTAGGAGATTAAGCTC 780
 Db 1416 TATAACCTGCAACCTCTACTAGCTCATGCGAGAGCAAAAGAGCTAGGAGATTAAGCTC 1475
 QY 781 CATAGAGATTTGGGCAAGTTTGAAGAAAGATCTACAGATGAGAGAGAGAGATTAAGCTC 840
 Db 1476 CATAGAGATTTGGGCAAGTTTGAAGAAAGATCTACAGATGAGAGAGAGAGATTAAGCTC 1535
 QY 841 CTTGTGTGAGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTAAGCTC 900
 Db 1536 CTTGTGTGAGATTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATTAAGCTC 1595
 QY 901 ATTGTCTGGCCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
 Db 1596 ATTGTCTGGCCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1555
 QY 961 A----- 961
 Db 1656 AAGTGAATCTCTCTCTTACTTTTAATTTTCTCTGATTTCTACTGATCTTA 1715
 QY 962 -----AGTGGCTCGGATGGGCAACCAAGAGCTGTGAAATCTTGCACAA 1007
 Db 1716 GAAATGTTACTTGTATGAGCTGGGCAACCAAGAGCTGTGAAATCTTGCACAA 1775
 QY 1008 GTATGAGAGCTTTAGAGACAGCATTTCTTATGATGATGATGATGATGATGATGATGATGAT 1067
 Db 1776 GTATGAGAGCTTTAGAGACAGCATTTCTTATGATGATGATGATGATGATGATGATGATGAT 1835
 QY 1068 TCTGATTTGAT 1127
 Db 1836 TCTGATTTGAT 1895
 QY 1128 AGCTGAGATGGGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1187
 Db 1896 AGCTGAGATGGGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1955
 QY 1188 TGTTCGCAACTGTATGAT 1247
 Db 1956 TGTTCGCAACTGTATGAT 2015
 QY 1248 CTCTCA----- 1253
 Db 2016 CTCTCAAGGTTCTCTCCCAAGAAATTTGATATGATGATGATGATGATGATGATGATGATGAT 2075
 QY 1254 -----AGG 1256
 Db 2076 TTAAGTTTGTGTCGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2135

QY 1257 CAAAACAGGCTGAATTTGAGTTGAATCATACCAAGAGATGTTGTAAGAGCTGAG 1316
 Db 2136 CAAAACAGGCTGAATTTGAGTTGAATCATACCAAGAGATGTTGTAAGAGCTGAG 2195
 QY 1317 GCAGATCTCTGAGAGACAAATGAGAGCTGAACTTAATGAACAGCTTCAAAACAGAA 1376
 Db 2196 GCAGATCTCTGAGAGACAAATGAGAGCTGAACTTAATGAACAGCTTCAAAACAGAA 2255
 QY 1377 CAAGCAGCCAGAGGCTGTTGAGAAATCTCTGAGAAATTAATGAGAGAGAGAGAGAGAGAT 1436
 Db 2256 CAAGCAGCCAGAGGCTGTTGAGAAATCTCTGAGAAATTAATGAGAGAGAGAGAGAGAGAT 2315
 QY 1437 TGCAGAGATTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1496
 Db 2316 TGCAGAGATTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2375
 QY 1497 AG----- 1498
 Db 2376 AGAGGATGATTTTCTTGAAGAAATCAAAACCTTGAATTTTGAATTAACCTGATTTCA 2435
 QY 1499 -----AGATGATGACACGA 1514
 Db 2436 CATTTTGTATTAATTTGTCACAAAACAAACCTGATGATTTGATGATGATGATGATGATGAT 2495
 QY 1515 CAGATTTTCAATGATTTCAATCAACAGATCCATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
 Db 2496 CAGATTTTCAATGATTTCAATCAACAGATCCATGAAGAGAGAGAGAGAGAGAGAGAGAGAG 2555
 QY 1575 TTTTCAGATGTTGACAGAT 1634
 Db 2556 TTTTCAGATGTTGACAGAT 2615
 QY 1635 TTAATCCCTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
 Db 2616 TTAATCCCTTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2675
 QY 1666 ----- 1665
 Db 2676 CTGGGCTTTTGTGTTTCAAAACCTAAGATGATGATGATGATGATGATGATGATGATGATGAT 2735
 QY 1666 -AGAGCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1724
 Db 2736 CAGAGCTGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2795
 QY 1725 GGAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1784
 Db 2796 GGAAGAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2855
 QY 1785 GCATCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1844
 Db 2856 GCATCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2915
 QY 1845 GTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1878
 Db 2916 GTTCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2949

RESULT 6
 AB025633 81365 bp DNA linear PLN 27-DEC-2000

LOCUS
 DEFINITION
 AB025633 BA000015
 ACCESSION
 VERSION
 AB025633.2 GI:10178221

KEYWORDS
 SOURCE
 ORGANISM
 Arabidopsis thaliana (thale cress)

Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE
 1 (sites)
 Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asamizu, E., Kotani, H.
 and Tabata, S.
 Structural analysis of Arabidopsis thaliana chromosome 5. X.

Sequence features of the regions of 3,076,755 bp covered by sixty
P1 and TAC clones
DNA Res. 7 (1), 31-63 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 81365)
Nakamura, Y.
Direct Submision
Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
Institute, Department of Plant Gene Research, 1532-3, Yana,
Kisarazu, Chiba 292-0812, Japan (E-mail: ynakam@kazusa.or.jp,
Tel:81-438-52-3935, Fax:81-438-52-3934)
On Sep 15, 2000 this sequence version replaced gi:4589439.
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/agd_graph.cgi?c=MOM1
Genes with similarity to proteins in the databases are described in
'product', or 'note' qualifiers. Genes that have no significant
protein similarity are described as 'unknown protein'.
The software programs used to predict genes include: Graal
(Informatics Group, Oak Ridge National Laboratory,
http://combio.ornl.gov/Graal-1.3/),
GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),
NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and
SplicePredictor (Volker Brendel, Stanford University,
http://gremml.zool.iastate.edu/cgi-bin/sp.cgi).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
http://genome.wustl.edu/eddy/tRNAscan-SE/).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K19M13 and the 3' clone is MRO11.

COMMENT

FEATURES

source

CDS

CDS

CDS

1. 81365
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone_1fb="M1coul P1"
/join(3184..3403,3474..3595,3681..3723,3777..3898,
4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)
/note="gb|AAF01580.1
gene_id:MOM1.24
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97251.1"
/db_xref="GI:8809710"
/translation="MHVQPGTSVDSPPAHKYDETTPPKVQFLERISTKVVDNAQNP
YGSADDDSSSPILSPVFEPPSSFSFGDDPLPGIEDLQISGEPPGHELOAG
YSINGTSCNFEWVCHLEBDGSVNYIDESRERTSNAGKQPNVLTADVDELYLAIEG
OPLDRNRKKRFAFGKYRDLTVLLDVHTLLDCSFSVEFFMFAIFI CYFLISG
ACEDDMQSNIEKLTGHTGASYSKSLAVGPVIMWAATLSIKREGVSIKCSIDLTAIE
KSASTYTPRGPALVLIIGSDSHSRADNSPDLISGRDIEVLTRIFTRVM
FPKALNSFYHFRCSCHKSRNT"
join(6317..6428,6508..6629,6714..6756,6841..6932,
7121..7169,7396..7466,7550..7663,7785..7882,8112..8181)
/note="gb|AAF01580.1
gene_id:MOM1.23
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97250.1"
/db_xref="GI:8809709"
/translation="MOPGTSVSPALNOADPTPISTKVVDNKTIFPDGDDAPL
PALELOISGEPPGHELOAGVSYNGTSCNFEWVCHLEDSVNYIGAKKPNVLT
ADVGLCLATIEVQPLDRNRKSELVYRANRDKTACHPENQSNIDKLTGHTGASYSK
SLAIGFVHMEATLSIEREGYTIKNDLITTEKFSASTAVKIPPEKPAELVLIIGSD
GSEHCLRVNDENPDIRLCKEERKVLRLVYVVFEOKL"
complement(join(8642..8705,8794..8863,8978..9387,

CDS

CDS

CDS

CDS

9584..10347))
/note="gene_id:MOM1.22
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97249.1"
/db_xref="GI:8809708"
/translation="MSMMKGSXGWTAFDLKORQGLSEVEGDPFPVSTSVNASF
GVRRGLRRNHSEPSFSSVLLPSPRPALITENKOCQNGERGGCCRRKEDTLSPVNS
HDLATYKIKENNSMADNDLIDVLLSTEDDPEMALAFKQWSSGCKEDDEPTKIEG
SSDNRSERYTRTEKVTSSVKMAASTEDACKTLENSDSSFLVNASDNKFPDDI
SELDIILORLOSIPLEPMEEDLYLSRKQALKYRASASNSHRAQNAFOYDHNAA
KOHSDKAREBDMLAAKLAKEAAKTIIGITNKONDITWKLDELHGLHATAVQALQERLOM
IEGHFTVRSVSPNRRGSKNALRASQSEPFRLDEEGMHCORTSREIRNLSLOVITG
IGKHSRGASLPLAVKTFEEDNRYRPFDETRPCVITVRPFRHS"
complement(12417..13424)
/note="contains similarity to unknown protein
gb|AAF27018.1
gene_id:MOM1.21"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97248.1"
/db_xref="GI:8809707"
/translation="MATDSQPNQKLTLPKTRIALTVISTMTDNAQRPDGTINRPLR
LDFRPAIPNPKPVNIVSTSDVFDQSDILMTFLTPYHSGDKIPIVVEFGGFAFLS
PNAYVDPNCRFPARKPAVYISVNYRLAPERRVAYQDDGALKYIEHNGSILPA
NADISRCPPAGDSAGGANTAHVATIRICRPSRSPFAVXVIGLISIQPFGEERTAE
KOLVGAFLVSPRTDMCKMAGLRNDHEAVNVGGRNADISLDPETRVVYVAGDPPL
KDWQSYVEWKLCKRKATLLEYPNMFAPYIFPLPEAGQILMKIKQVDNRVASLS
A"
join(14922..14981,15072..15167,15563..15666,15755..15891,
16096..16178)
/note="gene_id:MOM1.20"
/codon_start=1
/evidence=not experimental
/product="50S ribosomal protein L24"
/protein_id="BAA97247.1"
/db_xref="GI:8809706"
/translation="MGKRAEKLIPMKILNGDNVMIIRGDKGEGTITKRYRQNR
VIVEGKILKHKIKGPDHEGIFVFEAPLHASNVQVDPYTRGCRKGVKYLEBGT
VRVARGTSGSIIIRPEILIKIRATPRFTGPKOTPEFWEEQYTDATGSGMDPL"
join(16823..16952,17143..17320,17412..17453,17675..17851,
17943..18110,18199..18390)
/note="gene_id:MOM1.19"
/codon_start=1
/evidence=not experimental
/product="26S proteasome, non-ATPase regulatory subunit"
/protein_id="BAA97246.1"
/db_xref="GI:8809705"
/translation="MERLORLFGAGGGLGHASPDSPTLDSROYTISLALLKMLKHG
RAGVMEVGMLEAGFDEYTVRVVDVAMQSGVGEVENDHAPOTRMILMLKQIG
RPEVYNGYHSHFGCMLSGVDINTQGSFELANORAVAVVDPIQSXKGVYIDAFR
SINPTIMLGOEPRTNLTGHLNPSIQALIHGNRHYSAIINRYRNEERKMLN
LHKKTLGTLRRDTSKTNIEQVQMLSLAAGKYNVAVEDDELSEBKLAIVWGR
ODAKHGLBEHNSLMSNINVTQILMLDTTVV"
complement(join(18965..19043,19240..19298,19419..19536,
19666..19739,19908..19961,20387..20530))
emb|CAA22897.1
/note="contains similarity to unknown protein
gene_id:MOM1.18"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97245.1"
/db_xref="GI:8809704"
/translation="MDKNQOAIERKMLVGMVEDEDOQAADSESLSPMEDLNRCAL
TTKQFYGFALCLSGLTCTLLSMLVFNPVFGITFTLGNLMALGSTAFLGPPOROV
TMLDPAIRIVATVATVLIASIIIALFCALVVRNKLTLTLLIILEFTGLIYWSLSYIPFAR
TMVSKIEWTCDDTR"
join(12672..23634,23708..23999,24111..24355,24458..24624,
24715..24925)
/note="emb|CAB62356.1
gene_id:MOM1.17
similar to unknown protein"

```

/codon start=1
/evidence=not experimental
/protein_id="BAA97244.1"
/db_xref="gi:8809703"
/tranlation="MSSRAGPMSKERKUNVGGVRFVEVLQGLAGTLASSODPDGEG
EVIVSKNKKRPGNTSGKTWVSQNSPPRAKGGQOOGGSNVSGKGNVSRKGNQNG
IQANI GGRGRALSRKYDNNFVAPPPVPSRPBEGNNMRARGSKAHPAHPEDPVED
VDSASEENDSPALDDSDLLASDDYDSDVQSKSGSKQKMKRFRGSLDLSLIEO
INPOSEOMCPACONGPQGAIDWYNLHPILHARTGARGVRLKREKGLAEVLEKIDQMG
ASVTPGCEIYGVKQKGEDEKDEIYVWPVMTVIMTRLDKNDKMGLMGOBELLREYF
KDEBIALRAHVSPOGHRGMSVLMFESSATGLAEARHRELAEMGLDRLAMQKRM
FSGVRLQYLFPLATKODLIFPNHSGQKTRLEFKSYOEVMVVELNOISEDNQNLVY
FKNKLSKQNHAVLSESHLEIMSSEKTRPAEENRIVROTHQKOEKNEEMDADRF
MDSIKQIHERDRKAKEENFEMLQOQERAKVVGQOONINPSNDCKRAEVSIFEF
OEKEMEERFEVEREMLIKDCKEMDKKHHEHITDLEKPEALQELQMTGGLHND
D"
join(25614,.25838,26061,.26207,26324,.26438,26516,.26610,
26710,.26813,27522,.27765,27943,.28036,28115,.28305,
28385,.28604,28688,.28855,28944,.29092)
/notice="gb|AAC97420.1
gene_id="MOM1.16
strong similarity to unknown protein"
/codon start=1
/evidence=not experimental
/protein_id="BAA97243.1"
/db_xref="gi:8809702"
/tranlation="MAPPAGQTAAVAENAGADAPPQOORGFSGTISGIVRAVWFYF
ASKFSEKQKPMDDPTASQMLTPIPHKSGELDMWFLESEKFNDDGNRALYMRHTN
IYPVAWTPESIRKSLSLYSPETLQNNQSIYAHIPARSGEPIDPDPEYPLNFSR
THAVVTPPKOKRKKKSLSPSSKQSDSESPREYKQKQKDPKEVPEVMTSLMKN
VTINLVDPDQFADLVEPTTGNYPTIYFNEFWLBDKRTIPNTEVSLPLNTEISF
ISMNKKQFOQVDSQFQWRQSGMDLSEDEKLFVLEBNPLILLETIMFVMSHVF
DELAKRNOIQFNNKSMELGSAKSVLNPICQEVIFLYLLNDISTWMLTASGVC
IHFPIKQKAMRIEVDKSGMIPLRPFHDEBSYAKTKEXYDIAIKELSYLLLVGL
SYLSAIEHRHSKMSVWLLSLTSCVWYGFIMCPOPLIYKLSVAHILPMRWTKF
LNTIIDDLEAFVITKMPILHRLSLVRDVIPLIYQKVMVYPVDKTRVNERGFGDEBT
AAKTIITAEKEERDKTN"
complement(join(29439,.29576,29661,.29885,29985,.30268,
30341,.30493,30579,.30722,30956,.31484))
/notice="gene_id:MOM1.15
77.8%; Score 1462; DB 8; Length 81365;
Best Local Similarity 83.3%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 376; Gaps 4;

```

QY	421	GCAAGAGAGAGTTCTGCTCAGCA	CACAGCTGTGCAGAGATTTCTCAGCGGAGAGATGAT	480	
Db	23092	GCAAGAGAGAGTTCTGCTCAGCA	CACAGCGTGTGCAGAGATTTCTCAGCGGAGAGATGAT	23151	
QY	481	GTGATTAATGTTCTTGAGGAA	GAGAGATTCGATGCTTTGGATGATTTCTGATGACAC	540	
Db	23152	GTGATTAATGTTCTTGAGGAA	GAGAGATTCGATGCTTTGGATGATTTCTGATGAGAC	23211	
QY	541	CTTGCAAGTATGATTAATGAT	CTCGATGTGAGTCAAAAGAGCCATGATATCAGAAACGAG	600	
Db	23212	CTTGCAAGTATGATTAATGAT	CTCGATGTGAGTCAAAAGAGCCATGATATCAGAAACGAG	23271	
QY	601	AATATAGTGTTCAAAAAGTT	CTTTGGGCGAGTTGATAGCTTTGATCGATCGACAGATTAAT	660	
Db	23272	AATATAGTGTTCAAAAAGTT	CTTTGGGCGAGTTGATAGCTTTGATCGATCGACAGATTAAT	23331	
QY	661	GAACCAACAGAGCGAGTGC	ATGTCACGTTGTCAAGACGACCTGGGGCCATCGATGTTGG	720	
Db	23332	GAACCAACAGAGCGAGTGC	ATGTCACGTTGTCAAGACGACCTGGGGCCATCGATGTTGG	23391	
QY	721	TATTAACCTGCACCCCTCA	CTCTAGCTCATGCGAGGACAAAGAGAGCTAGCGAGTTAAGCTC	780	
Db	23392	TATTAACCTGCACCCCTCA	CTCTAGCTCATGCGAGGACAAAGAGAGCTAGCGAGTTAAGCTC	23451	
QY	781	CATAGAGAAATTTGGCTGA	AGTTTGAAGAGATCTACAGATGAGAGGCGCATCTGCATT	840	
Db	23452	CATAGAGAAATTTGGCTGA	AGTTTGAAGAGATCTACAGATGAGAGGCGCATCTGCATT	23511	
QY	841	CTTTGTGTGTGATTTTAT	GGGCGAGTGGAGAGGGTTTGGGTGAGAGATGAAAAAGATTATGA	900	
Db	23512	CTTTGTGTGTGATTTTAT	GGGCGAGTGGAGAGGGTTTGGGTGAGAGATGAAAAAGATTATGA	23571	
QY	901	ATTGTCGCGCTCCAAATG	GTGCATCATCATGAAATCTACAGCTGGATTAAGACATTAACGAT	960	
Db	23572	ATTGTCGCGCTCCAAATG	GTGCATCATCATGAAATCTACAGCTGGATTAAGACATTAACGAT	23631	
QY	961	A-----	-----	961	
Db	23632	AAGGTGAATCTTCTGTCTT	TAACTTTTCTCTTGATCTTACTGATCTTAA	23691	
QY	962	-----	AGTGGCTCGGATGCGCA	ACCAAGAGCTGCGAAATCTTCGACAA	1007
Db	23692	GAATGTTAATTTAGTGTG	CGCGGCAATGGGCGCAACCAAGAGCTGCGAAATCTTCGACAA	23751	
QY	1008	GTATGAGGCTTTAGAGAG	CGCCATTTCTATGTGCCACAGGGCCATCGTGGAGTGAAGTGT	1067	
Db	23752	GTATGAGGCTTTAGAGAG	CGCCATTTCTATGTGCCACAGGGCCATCGTGGAGTGAAGTGT	23811	
QY	1068	TCTGATGTTTGAAGAC	ATGTCACCTGCTGATATTTGAGAGCCGCAAGGCTCCACCGGAGATT	1127	
Db	23812	TCTGATGTTTGAAGAC	ATGTCACCTGCTGATATTTGAGAGCCGCAAGGCTCCACCGGAGATT	23871	
QY	1128	AGCTGAGATGGGTTAGAT	AGAAATGCTCGGGGTCAAGAGCGCAGATATGTTTTCGAGG	1187	
Db	23872	AGCTGAGATGGGTTAGAT	AGAAATGCTCGGGGTCAAGAGCGCAGATATGTTTTCGAGG	23931	
QY	1188	TGTTGCGCAACTGTATG	GCTTCTTGCAACGAGCAAGATCTGGAATATTAATCAATCAACA	1247	
Db	23932	TGTTGCGCAACTGTATG	GCTTCTTGCAACGAGCAAGATCTGGAATATTAATCAATCAACA	23991	
QY	1248	CTCTCA-----	-----	1253	
Db	23992	CTCTCAAGGTTCTCTCC	CCCAAGAAATTTGATATATGCTTTTAGTTTGTCAATGSAAT	24051	
QY	1254	-----	-----	1256	
Db	24052	TTAAAGTTTGTGTC	CGGTGAATGATCATGTTATGTATATATCATGATTAATTAAG	24111	
QY	1257	CAAAACAGGCTGAATTTG	AGTTGAATCATCAACAGAGATGGTTGAAGAGAGCTGAG	1316	
Db	24112	CAAAACAGGCTGAATTTG	AGTTGAATCATCAACAGAGATGGTTGAAGAGAGCTGAG	24171	
QY	1317	GAGATCTCTGAGGAC	AATCAGAGCTGAACCTTAATTAAGACAAAGCTCTCAAAACAGAA	1376	

```

Db      24172 GCAGATCTCGAGACAATCAGACGTGAACTATTAGAACAACTCTCAAAACAGAA 24231
      1377 CAAGCAGCGCAAGGCTTGAGGAATCTCTGGAATTTATGAGAGAGAGCTGGTGAAC 1436
      24232 CAAGCAGCGCAAGGCTTGAGGAATCTCTGGAATTTATGAGAGAGAGCTGGTGAAC 24291
      1437 TGCAGAGATTAATCGATCGTGAAGAGAACTAAGATGCAACATGAAACAGAGGGA 1496
      24292 TGCAGAGATTAATCGATCGTGAAGAGAACTAAGATGCAACATGAAACAGAGGGA 24351
      1497 AG----- 1498
      24352 AGAGGATGATTTTCTCTAGAAAATCAAACTTGACATTGTATTAAGTCTAGATTC 24411
      1499 -----AAGTGAATGCACACCA 1514
      24412 CATTTTGTATTAATTTGTCACAAAAAACCTGTGTGTTGTAAGATGATGCACACGA 24471
      1515 CAGGTTTTTCATGATTCATCAATCAACAGATCCATGAAAGAGAGAGAGAGAGAA 1574
      24472 CAGGTTTTTCATGATTCATCAACAGATCCATGAAAGAGAGAGAGAGAGAGAA 24531
      1575 TTTTCAGATGTTGACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1634
      24532 TTTTCAGATGTTGACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 24591
      1635 TAAATCCCTTAGAATGAGATGTCGGAAG----- 1665
      24592 TAAATCCCTTAGAATGAGATGTCGGAAGAGATGATGACTAACTAATATATCCCT 24651
      1666 ----- 1665
      24652 CTGGGCTTTTGTGTTTCAAACTAAGATGATGATTAATCCGCTTTGATCTTTGG 24711
      1666 -AGAGCTGAGAGAGTGTCAAGCTTCATCGATTTCAAGAGAGAGAGAGAGAGAGAG 1724
      24712 CAGAGCTGAGAGAGTGTCAAGCTTCATCGATTTCAAGAGAGAGAGAGAGAGAGAG 24771
      1725 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1784
      24772 GGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 24831
      1785 GCATCAGAGAGAGATTTGATCTGAGAGAGAGATTTGATGAGAGAGAGAGAGAGAG 1844
      24832 GCATCAGAGAGAGATTTGATCTGAGAGAGAGAGATTTGATGAGAGAGAGAGAGAG 24891
      1845 GTACAGAGAGAGAGATTTGATCTGAGAGAGAGAGATTTGATGAGAGAGAGAGAGAG 1878
      24892 GTACAGAGAGAGAGATTTGATCTGAGAGAGAGAGATTTGATGAGAGAGAGAGAGAG 24925

```

```

RESULT 7
ATHS28171 650 bp DNA linear PIN 29-MAR-2003
LOCUS Arabidopsis thaliana T-DNA flanking sequence, left border, clone
DEFINITION 157608
ACCESSION AJ528171
VERSION AJ528171.1 GI:26796431
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
REFERENCE
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F.,
Chavuin, S., Rechold, N., Cruaud, C., DeRose, R., Pelletier, G.,
Lepoint, L., Caboche, M. and Lecharny, A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences
of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
MEDLINE 22363535

```

```

PUBMED 12446565
REFERENCE 2 (bases 1 to 650)
AUTHORS Balzerque, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue
Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT
PCR was performed on DNA from transformants of Arabidopsis thaliana
plants from INRA (Versailles). The DNA fragment(s) resulting from
the PCR were directly sequenced from the left or the right border
to determine the genomic sequence flanking the insertion. T-DNA
derived sequences were removed. Information to order the
corresponding mutant line and a link to a database providing a
graphical display of the insertion site are available at
http://dbsgap.versailles.inra.fr/publiclines/. This sequence has
been generated in the framework of the French plant genomics
program 'Genoplante' (http://www.genoplante.com and
http://genoplante-info.infobiogen.fr).
FEATURES
source
1..650
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="nasbillewskija"
/db_xref="taxon:3702"
/clone="157608"
/clone_1ib="Arabidopsis thaliana T-DNA insertion lines"
misc_feature 1..650
/note="T-DNA flanking sequence
left border"
ORIGIN
Query Match 23.3%; Score 437.8; DB 8; Length 650;
Best Local Similarity 87.4%; Pred. No. 96-96;
Matches 522; Conservative 0; Mismatches 2; Indels 73; Gaps 1;
QY 733 CCTTACTAGCTCATGCGAGACAAAGAGAGCTAGCGAGTTAAGCTCATAGAGAAATTG 792
DB 1 CCTTACTAGCTCATGCGAGACAAAGAGAGCTAGCGAGTTAAGCTCATAGAGAAATTG 60
QY 793 GCTGAAGTTTGAAGAGATCTACAGATGAGAGGCGATCGTCACTTCTGTGTGAG 852
DB 61 GCTGAAGTTTGAAGAGATCTACAGATGAGAGGCGATCGTCACTTCTGTGTGAG 120
QY 853 ATTATGCGCAGTGAAGAGGTTGGTGTGAGATGAAGAAAGATTATGAATTTCTGGCCT 912
DB 121 ATTATGCGCAGTGAAGAGGTTGGTGTGAGATGAAGAAAGATTATGAATTTCTGGCCT 180
QY 913 CCAATGTCATCATGATGAATACTAGCTGATGAAGAGATTAAGAGATTAAGAGATTA 961
DB 181 CCAATGTCATCATGATGAATACTAGCTGATGAAGAGATTAAGAGATTAAGAGATTA 240
QY 962 ----- 961
DB 241 TTCTGCTTTTAACTCTTTAATTTTCTCTTGCAATTCATGATCTTGAATGTTAATT 300
QY 962 --AGTGGCTCGGATGAGGACCAAGAGCTGTGGAATCTTCAAGATGAGGCTCT 1019
DB 301 GTATGTCCTGCGCATGCGGACCAAGAGCTGTGGAATCTTCAAGATGAGGCTCT 360
QY 1020 TAGAGCAGCCATTCCTATGTCACAGGAGCATGCTGAGATGAGTGTCTGATGTTGA 1079
DB 361 TAGAGCAGCCATTCCTATGTCACAGGAGCATGCTGAGATGAGTGTCTGATGTTGA 420
QY 1080 GAGCAGTGCATGCTGATTTGAGAGCCGAAGCCCTCCACGGGAGTTAGCTGAGATGGG 1139
DB 421 GAGCAGTGCATGCTGATTTGAGAGCCGAAGCCCTCCACGGGAGTTAGCTGAGATGGG 480
QY 1140 GTTATGATGAATTTGCTGGGAGTCAAGAGGAGTATGTTTCTGAGAGGTTGCGCAACT 1199
DB 481 GTTATGATGAATTTGCTGGGAGTCAAGAGGAGTATGTTTCTGAGAGGTTGCGCAACT 540
QY 1200 GTATGCTTCTCTTGAACAGAGAGATCTGAGATTAATTAATCAACTCTGAAGG 1256
DB 541 GTATGCTTCTCTTGAACAGAGAGATCTGAGATTAATTAATCAACTCTGAAGG 597

```


,

Db

1010 TATTCAGTGAATGTGCAAGCAGACGCCAIGCCIAIAGTCTCGAAAGGGCGGG- -

Oy 1057 GGGATGAGTGTCTGATGTTTGAAGACAGTGCCTGCTATTGGAGGCCGACCCCTC 1116
 |||||
 Db 1070 GGCAATGAGTGTCTATCTTTGACAGCTGGCTGTGGCTATATGAGACAGACCTTTG 1129
 |||||
 Oy 1117 CACCGGAGTACTGATGATGGGCTTATGATGATCTCTGGCT-----CAGAGCG 1169
 |||||
 Db 1130 CATGATCATCTTGTGTCTGTAAGAGACAGAGATATGAGAACAGTCTCCACCAAGT 1189
 |||||
 Oy 1170 CAGTATGTTTCTGAGAGTGTTCGCCAATGATGCTCTCTTGCACGACGACGATCT 1229
 |||||
 Db 1190 TACATCTTACCTGCTGGGAGAGAGAGCTGATGCTTTCTTACCTAACAAAGATACAT 1249
 |||||
 Oy 1230 GGACATATTCATCAACACTCTCAAGGCAAAACAAAGCTGAAATTCGAGTTGAAATCAT 1289
 |||||
 Db 1250 GGAGACATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1309
 |||||
 Oy 1290 CCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1349
 |||||
 Db 1310 TAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369
 |||||
 Oy 1350 CTTTAAAGAACAGCTCTCAAAACAGACAGACAGACAGACAGACAGACAGACAGACAG 1409
 |||||
 Db 1370 CCGTGAAGAACAAATGCTTAAAGAGACAGACAGACAGACAGACAGACAGACAGACAG 1429
 |||||
 Oy 1410 AATTATGAGCAGAGAGCTGCTAGAACTGACAGAGATATTCGATGCTGAGACAGAGAAC 1469
 |||||
 Db 1430 TCTTTTACCCAGAACTTAAAGAGACAGATGAGAGAGATATCTTTTACGAAATTAAC 1489
 |||||
 Oy 1470 TAAGATGACAGATGAAACAGACAGAGAGATGATGACAGACAGAGCTTTTTCATGGA 1529
 |||||
 Db 1490 CAAAGAAAAAGCACTTGAGATATGAGAAAGATGAAATGACAGAGAGAAATTTTTCATGA 1549
 |||||
 Oy 1530 TTCAATC-AAAGATCCATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1588
 |||||
 Db 1550 TCAATTCGAAAAAATTCACAAAGCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1609
 |||||
 Oy 1589 AGCAGCAG 1648
 |||||
 Db 1610 TGCAG 1660
 |||||
 Oy 1649 ATGACGATGTCGAAAG 1708
 |||||
 Db 1661 CTGAAGATGACAG 1720
 |||||
 Oy 1709 AGATGAG 1768
 |||||
 Db 1721 ATGTTGAG 1780
 |||||
 Oy 1769 AAGACATGAAAG 1828
 |||||
 Db 1781 TGAAGCTCAAG 1840
 |||||
 Oy 1829 CTTTGAAGACGCTCATGTAACAGCA 1853
 |||||
 Db 1841 CCTCAGCTTCCCTGATGATTAACA 1865
 |||||

RESULT 9
 AKI00699 2570 bp mRNA linear PLN 24-JUL-2003
 LOCUS AKI00699
 DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone: j023114M02, full
 insert sequence.
 ACCESSION AKI00699
 VERSION AKI00699.1 GI:32985908
 KEYWORDS FLI CDNA; CAP tripper.
 SOURCE Oryza sativa (japonica cultivar-group)
 ORGANISM Oryza sativa (japonica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Eriatridae; Oryzaceae; Oryza.
 REFERENCE
 AUTHORS The Rice Full-Length cDNA Consortium, National Institute of

COMMENT

TITLE
JOURNALREFERENCE
JOURNAL
MEDLINE
PUBMED
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMEDREFERENCE
JOURNAL
MEDLINE
PUBMED
AUTHORS

Agrobiological Sciences Rice Full-length cDNA Project Team:
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
 Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
 Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group: Otsu, Y., Murakami, K.,
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
 Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
 Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nishikawa, J.,
 Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
 Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
 Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
 Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
 Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Otsu, N., Ota, Y.,
 Satoh, R., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T.,
 Yoshino, M., and Hayashizaki, Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from
 japonica rice
 Science 301 (5631), 376-379 (2003)
 22752273
 12869764
 2 (bases 1 to 2570)
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
 Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
 Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
 Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
 Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
 Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
 Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Kurotsaki, T., Kusumegi, T., Li, C., Lu, M.,
 Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
 Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
 Namiki, T., Narikawa, R., Nishikawa, J., Nishikawa, K., Nomura, K.,
 Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
 Osato, N., Ota, Y., Otsu, Y., Ryu, R., Satoh, H., Sakai, K., Sakai, K.,
 Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
 Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
 Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
 Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
 Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
 Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S., and
 Yoshimura, A.
 Direct Submission
 Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
 Agrobiological Sciences, Department of Molecular Genetics, Head of
 Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
 305-8602, Japan (E-mail: skikuchi@niae.affrc.go.jp,
 Tel:81-29-838-7007, Fax:81-29-838-7007)
 This clone is one of the 28K full-length cDNA clones from japonica
 rice. URL: <http://cdna01.dna.affrc.go.jp/cDNA/>
 NIAS Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K.,
 Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,
 Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,
 Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., and
 Yamamoto, M.
 Pairs Genome Sequencing & Analysis Group: Otsu, Y., Iida, Y.,
 Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
 Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
 Mizuno, K., Narikawa, R., Nishikawa, J., Oka, M., Ryu, R., Sugano, S.,
 Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
 Yoshimura, A., Matsubara, K., and Murakami, K.
 Genome Exploration Research Group in Riken Genomic Sciences Center
 and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
 Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
 Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
 Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
 Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
 Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N.,
 Ota, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,

REFERENCE
2 (bases 1 to 1556)
Aadachi,J., Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
Fujimura,T., Fukuda,S., Hanagaki,T., Hara,A., Heshizume,W.,
Hayashida,K., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
Hori,F., Hotei,I., Iida,J., Iida,Y., Ikeda,R., Imanura,K.,
Imocani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
Kawagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,H.,
Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koye,S., Kurihara,C., Kurotsuki,T., Kusumegi,T., Li,C., Lu,M.,
Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
Nami,K., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K.,
Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
Osato,N., Ota,Y., Otsu,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,M.,
Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
Yoshimura,A.

TITLE
Journal
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 28K full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-Length cDNA Project Team: Kikuchi,S., Satoh,K.,
Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
Ishikawa,M., Yamada,H., Ooka,H., Hotei,I., Kojima,K., Nami,K.,
Ohneda,E., Yahagi,M., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
Yamamoto,M.

FAIS Genome Sequencing & Analysis Group: Otsu,Y., Iida,Y.,
Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
Kodama,T., Kurotsuki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
Mizuno,K., Narikawa,R., Nishikura,J., Oka,M., Ryu,R., Sugano,S.,
Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
Yoshimura,A., Matsubara,K. and Murakami,K.

Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken:
Aadachi,J., Aizawa,K., Carninci,P., Fukuda,S., Hanagaki,T.,
Hara,A., Heshizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
Hiraoka,T., Hori,F., Iida,J., Imanura,K., Imocani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koye,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,
Saeki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
Yasunishi,A. and Hayashizaki,Y.

FEATURES
source
1. 1556
Location/Qualifiers
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="002-104-F07"

ORIGIN
Query Match 19.6%; Score 368; DB 8; Length 1556;
Best Local Similarity 60.7%; Pred. No. 9,6e-79;
Matches 659; Conservative 0; Mismatches 415; Indels 12; Gaps 3;
460 TTCTCTGAGTGGAGATATGATATGCTTTCTGAGGAAGAGATATTCGATGCT 519
473 TTGCTGTTGATGATCCGAAACAGACATGCTGAAGGTTATGACATCGACGATGAT 532

QY 520 TTGATGATTTCTGATGACGACCTTGCAAGTGAATTATGACTCGATGATGACAAAG 579
DB 533 AATGATGATGATATGATGACGATTTAGTGTGACTATGATTTCTGATGATGAGAA 592
QY 580 ACCCATGATGATCGAAGAGAAATAGTGTTCAAAAGTTCTTTGGACGTTGATGAC 639
DB 593 AATTGATGATCGGAAATATCAAGATTTTCAAGGTTCTTTCGAACTCTTGAGGGC 652
QY 640 TTGTCGATGACGACGATTAATGAACAACAGAGAGTGGCATTTGTCACCTTGAGAC 699
DB 653 TTAAGTGTGAACAGTAAATATGAACACTAGGCAATGGCATTTGCCAGCATGCAAAAT 712
QY 700 GGACCTGTGGCCATGATGATTAATAA---CTGCAACCTCTACTAGCTCATGCGAGACA 756
DB 713 GGACCTGTGGCAATGATGATGATTAACAAAGGGCTCAACCTTATGATGATGATGCA 772
QY 751 AAAGAGCTGAGCGAGTTAAGTCCATGAGAAATTTGCTGAAGTTTGAAGAAATCTTA 816
DB 773 AAGGCTTCTATTAAGTCAAGGCTCAAGAAATTTGCTAGTTTATGGAAGAGCTTA 832
QY 817 CAGATGAGAGGCGCATCTGTCAATTCCTGTGTGAGATTTATGGCAGTGAAGGTTTG 876
DB 833 TTCTCGAGGGGAATTCAGTGTACCATCTGCGCAATTTAGAAATGGAAGGATTTG 892
QY 877 GGTGAGATGAAAGATTAATTAATTTGCTGCTCCATGATGATCATCATGATTAAT 936
DB 893 CG---GGAAGCATGATGATGATTTGTTGGCCAAATGTTGTTTATGAACAT 949
QY 937 AGACTGATTAAGACGATTAACGATTAAGTGTCTGCGATGCGCAACCAAGCTGTGGA 996
DB 950 GTATTGAAACAGATGAGATGATTAAGTGAAGGTAATGGGAACCAAGACTTAATGAT 1009
QY 997 TACTTCGACAGATGAGGCTTTAGACAGCCATTCCTATGATGATGATGATGATGAT 1056
DB 1010 TATTTCACTGAATTAATGCTGCAAGCAAGCCTGCTTAATGATGATGATGATGAT 1069
QY 1057 GGGATGAGTGTCTGATGTTTGAAGAGAGTGCATGCTATTTGAGGCGCAAGCCTTC 1116
DB 1070 GGCATGAGTGTCTGATGATTTTGAAGAGTGCATGCTATTTGAGGCGCAAGCCTTC 1129
QY 1117 CACCGGAGTTAGCTGAGATGAGGTTTGAATGATGATGATGATGATGATGATGAT 1170
DB 1130 CATGATCACTTTGCTGCAAAAGCAAGCAAGATTAATGATGATGATGATGATGAT 1189
QY 1171 AGTATGTTTTCGAGTGTTCGCAACGATGATGATGATGATGATGATGATGATGAT 1230
DB 1190 ACATTTCTTACCGGTGGAAGAAAGCAAGCTGATGATGATGATGATGATGATGAT 1249
QY 1231 GACATATTCAATCAACCTTCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1290
DB 1250 GAGCATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1309
QY 1291 CAGAGATGTTTGAAGAGAGTGAAGAGATCTTGAAGCAATGATGATGATGATGATGAT 1350
DB 1310 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1369
QY 1351 TTTAAGCAAGCTTCAAAACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1410
DB 1370 CTGAAGAAACAAATGTTTGAAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1429
QY 1411 ATTATGAGAGAGAGTGTGAAGCTGAGAGATTAATGAGATGATGATGATGATGATGAT 1470
DB 1430 GTTGTATCCCAAACTTAAGGAGCAATGAGAGAAATATGATGATGATGATGATGAT 1489
QY 1471 AAGATGAGATGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1530
DB 1490 AAAGAAAGCACTTGAATGAGAAAGATGAAGTACAGAGAGAAATTTTCATGAT 1549
QY 1531 TCAATC 1536
DB 1550 CAATC 1555

RESULT 11
 AF542974 1564 bp mRNA linear PLN 02-JUL-2003
 LOCUS Triticum aestivum Em1 mRNA, complete cds.
 DEFINITION
 AF542974
 ACCESSION
 AF542974.1 GI:32401385
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Triticum aestivum (bread wheat)
 Triticum aestivum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Poideae; Triticeae; Triticum.
 1 (bases 1 to 1564)
 Zhao, X., Li, Q. and Zhang, X.
 Isolation and expression of a new kind of gene involve in
 embryogenesis in Triticum aestivum L.
 REFERENCE
 2 (bases 1 to 1564)
 Unpublished
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Submitted (03-SEP-2002) College of Life Sciences, Shandong
 Agricultural University, Daizong Street, Tai'an, Shandong 271018,
 P.R. China
 Location/Qualifiers
 1..1564
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /cultiyar="PH 82-2-2"
 /db_xref="taxon:4565"
 /dev_stage="embryo"
 309..1223
 /note="involved in embryogenesis"
 /codon_start=1
 /product="Em1"
 /protein_id="AAP0862.1"
 /db_xref="GI:32401386"
 /translation="MGNOELLEFPDYATKAAHAYPGHGRMSVLIPESSAVGYME
 ARLHKEPIIDORTDQNRNRPFLPGKRLQGLFARKEDMETRRHRCQGSRLKY
 EMRSNENWVAOMKMSNDQNLNLYLNKLVKTKORSKVEEFTLVYTKLRRTMEN
 IFRSKAKEXSEVEEMKSGKEFHQIENIHATDKSEFEERLQGERARQCD
 VDSGTENRRLRKEVOYRPFICGVKMDQVEFAEDEMI KAHKEKVKQLKKEYAKAVE
 LEKFPDALTLMEKRPRTQASSSSP"
 ORIGIN
 Query Match 15.5%; Score 291.6; DB 8; Length 1564;
 Best Local Similarity 59.2%; Pred. No. 4.2e-60;
 Matches 539; Conservative 0; Mismatches 359; Indels 12; Gaps 2;
 Oy 952 GATACGATTAAGTGGCTGGCATGGGCAACCAAGAGCTGCGATTAAGTGAAGTAT 1011
 Db 288 GATGCTCAGTGGTGAAGGCAAGGCAACCAAGACTTCTGATTAATTTGATGATAT 347
 Oy 1012 GAGGCTTTAGAGAGCGCATTCCTATGTGTCACAGGCGCATGCGGATGATGTTCTG 1071
 Db 348 GCTGCGAACAAGACCGCATGATATGTCAGTGGGCAACCGTGGCATGATGTCGA 407
 Oy 1072 ATGTTTGAAGCACTGCTGCTGCTATTTGAGAGCGGAGCGCTCCACGGAGTAACT 1131
 Db 408 ATATTGAAAGCTGCTGCTGCTGCTATATGAGAGCGGAGCGCTTCTTCAATGAAGATTAT 467
 Oy 1132 GAGATGGGTTAGATAGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1188
 Db 468 GATCAAAAGACAGACAGGACACTTGGCAAGATGCGAGGCTTCTTCACTGCTGCTGCTG 527
 Oy 1189 GTTCCCAACTGATGCTTCTTCTTCAACGAAAGCAATCTGACATATTCATCAACAC 1248
 Db 528 AAGAGACATTAATACGCTTCTTCTTCAACGAAAGCAATGAGATTTCAATTAACAC 587
 Oy 1249 TCTCAAGCAAAAGGCTGAAATTCGATTTGAATCATACCAAGAGATGTTGAAG 1308
 Db 588 TCCCAAGGAGAAAGCCCTGAAATACAGATGATCTCATATGAGATGATGCGCC 647
 Oy 1309 GAGTGAAGGCAATCTCTGAGGACATGACAGCTGAACCTTGAAGCAAGCTCTCA 1368

Db 648 CAGATGAAGCAAAATGACCGAAGTAAATCAACAACTCAATTAATCTGAAGAAACAGGCTGTT 707
 Oy 1369 AAACAGAAACAGACAGCGCAAGTGTCTTGAAGATCTCTGAAATTTGAGCGGAGAACTG 1428
 Db 708 AAGACAGACAGCGCTCTAAAGTTGTAGAAAGAAACCTGGGTGTATTTACAGAAACTT 767
 Oy 1429 CGTAGAATTCAGAGATTAATCGATTCGTGAGACAGAACTAAGATGACATGAACAG 1488
 Db 768 CGAGAGCTATGAGAAAGAAATATTTCTGTCAGAGATTAACCTAAAGAGCACTCTGAG 827
 Oy 1489 AACAGAGAGATGATGACACAGACAGTTTTCATGGAATTCATCAACAGATCCAT 1548
 Db 828 TATGAGAGAGATGAATCCACAGAAATTTCTTCATATGATGATGAGAACATTCAC 887
 Oy 1549 GAAAGAGAGACGCAAGAGAGAAATTTGAGATGTCAGACAGCAAGAGTCCAG 1608
 Db 888 AAGGCCACAGAAAGACAGAAAGCGAGTTGAGATTCCTGAGAGAGAGCGTGAAG 947
 Oy 1609 GTTGTGGCAGACAGACAGAAACATTAATCCCTTACAAATGACATTTGCCGAAAGAG 1668
 Db 948 GCT-----AGACAGTGTATGATGATCTGAGACTGAAATGCGAGCTTAAG 998
 Oy 1669 GCTGAGAAAGTGTCAAGCTTCTTCAAGTTCAAGAAAGATGAGAGATTTGTGAA 1728
 Db 999 AAGAACAGGTACAGCGGTTCTATGAGTCCAGTTAAGACCTGACAGAGTTGAGGCT 1058
 Oy 1729 GAGAGAGATGCTGATTAAGATCAAGAAAGAAAGATGAAAGACATGAAGAGGAT 1788
 Db 1059 GAAAGGACCAATGATTAAGAGCGCAGAGAGAAAGATGACACTGAAGAAAGATGAC 1118
 Oy 1789 CACGAGAGATATTGATCTGAGAAAGAAATTTGATGAGGCTTTGAAACAGCTGATAC 1848
 Db 1119 ATGCGAAAGAGAGGAGCTCGAAGAGAGTTGATGCTGCTCAACCGTTGATGAG 1178
 Oy 1849 AAGCATGGCC 1858
 Db 1179 AAACATGAGC 1188
 RESULT 12
 AK061975 1214 bp mRNA linear PLN 24-JUL-2003
 LOCUS Oryza sativa (japonica cultivar-group) cDNA clone:001-043-A02, full
 DEFINITION
 AK061975
 INSERT SEQUENCE
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 The Rice Full-Length cDNA Consortium, National Institute of
 Agricultural Sciences Rice Full-Length cDNA Project Team;
 Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
 Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
 Kojima, K., Nanki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
 Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
 Science Genome Sequencing & Analysis Group; Otsu, Y., Murakami, K.,
 Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
 Kurosaki, T., Kodama, T., Maeda, H., Kobayashi, M., Xie, Q., Lu, M.,
 Nishikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Miura, J.,
 Iweda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., KIKEN,
 Kuwamegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN,
 Kawai, J., Carninci, P., Adachi, U., Aizawa, K., Arikawa, T., Fukuda, S.,
 Hara, A., Hashidume, W., Hayatsu, N., Imetani, K., Ishii, Y., Itoh, M.,
 Kagawa, I., Kondo, S., Komori, H., Miyazaki, A., Otsu, N., Ota, Y.,
 Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
 Yoshino, M., and Hayashizaki, Y.
 Collection, mapping, and annotation of over 28,000 cDNA clones from
 Japonica rice

JOURNAL Science 301 (5631), 376-379 (2003)
MEDLINE 22752273
PUBMED 12869764
REFERENCE 2 (bases 1 to 1214)
AUTHORS Adachi, J., Aizawa, K., Akiyama, T., Arikawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imocani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawana, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyu, S., Kurihara, C., Kurose, T., Kusumegi, T., Li, C., Lu, M., Masuda, M., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, M., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)
COMMENT This clone is one of the 28k full-length cDNA clones from japonica rice.
URL: http://cdna01.dna.affrc.go.jp/cDNA/
NIA Rice Full-length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, M., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.
PALS Genome Sequencing & Analysis Group: Otomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawana, M., Kobayashi, M., Kodama, T., Kurose, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nishikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akiyama, T., Arikawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Imamura, K., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, C., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koyu, S., Kurihara, C., Kurose, T., Kusumegi, T., Li, C., Lu, M., Masuda, M., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nishikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.
FEATURES
source location/Qualifiers
1..1214
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cufv="Nipponbare"
/db_xref="taxon:39947"
/clone="001-043-A02"
ORIGIN
Query Match 15.4%; Score 288.8; DB 8; Length 1214;
Best Local Similarity 58.7%; Pred. No. 2e-59;
Matches 544; Conservative 0; Mismatches 367; Indels 15; Gaps 2;
934 ACTAGACTGATAGAGACGATTAAGTGTCTCGGATGCGCAACCAAGAGCTGCTG 993

Db 2 ACTGATTTGGAACAAGATGAGGATGATTAAGTGGAAAGGATATGCGGAAACCAAGAGCTAAAT 61
994 GAATCTTCGACAAAGATATGAGGCTCTTGAGACGCGATTCCTATATGCTCAAGGCGCAT 1053
Db 62 GATTTATTTAGTGAATATGCTCAAGCAAGCAAGCAGCGATGCTATGCTCAATATGGCGAC 121
1054 CCTGGATGAGTGTCTGATGTTGATGTTGAGACAGTGCACCTGCTATTTGGAGGCCGCAACGC 1113
Db 122 CCGTGCATGATGCTGTATCTTTGACAGCTCGGCTGTATATGAGAGAGAACT 181
1114 CTCACCGGAGTATGCTGAGATGGGTTAGATGAATGCTCG-----GCTCAAG 1167
182 TTGCATGATCACTTTGTTGCTCAAGAAACAGACAGAAATACATGCAACAGTCTTCAAG 241
1168 CCCAGTATCTTTCTGAGAGTGTGGCCAACTGTATGCTTCTTGCAACGAAAGAT 1227
242 GTTACATTTCTTACCTGTGGGAAAGGCGCTGTATGTTTCTTAGCTCAAAAGATGAC 301
1228 CTGACATATTCATCACTCTCAAGGCAAAACAAAGGCTGAATTCAGTTGAATCA 1287
302 ATGGAACATTCATATGACATTTGCCAGGAAATCCGTTGATATATGATGATGATCT 361
1288 TACCAAGAGATGTTGTAAGAGAGCTGAGCAGATCTTGAGACATCAGACGTTGAAC 1347
362 TATATGAGATGTTGTTTACCAATGAAAGCAGATGAGGAGAACACAGCACTGAAT 421
1348 TACTTAAAGAACAGCTCTCAAAACAGAACAGACGCGCAAGCTGTTAGAAATCTCG 1407
422 TACCTGAAGAAACAAATGTTTGAAGAGCGACCTTAAATTTGATAGAGATACATTTG 481
1408 GAATTAATGAGGAGAGAGCTGCTGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1467
482 ACTGTTGTTACCAAGAACTTGAAGAGCAATGAGAAAGAAATACATTTAGAAATTA 541
1468 ACTAATGACAGCATGAAACAGAAAGAGAGATGATGACACAGACAGGTTTTTCATG 1527
542 GCCAAAGAAAGCACTTGAGATGATGAAAGAGATGAAAGTACAGAGAGAAATTTTCCAT 601
1528 GATTCAATCAACATCTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 1587
602 GATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 661
1588 CAGCAGCAGAAAGCGCAAGGTTGTGGCCAGCAGCAGCAGCAATTAATCCCTCTAGC 1647
662 CTGCAAGAGAGCGAGCAAGGCT-----AAGCAAGTGAATTTGACTCTGGAAT 712
1648 AATGACATGTCGCAAG 1707
713 ACTGAAGATCGCAGGCAAG 772
1708 GAGATGAGAGAGATTTGTGAG 1767
773 GATGTTGAGAGATTTGAG 832
1768 GAAGCATGAG 1827
833 GTGAAGCTCAG 892
1828 GCTTGAAG 1853
893 GCCCTCAGCTTCCCTGATGATGATTAACA 918
RESULT 13
AF469493 513 bp mRNA linear PLN 02-JUL-2003
LOCUS Triticum aestivum H0TR mRNA, partial cds.
DEFINITION AF469493
ACCESSION AF469493.1 GI:32400755
KEYWORDS Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 513)
Li, J.R., Wang, F., Li, Q.Z. and Zhang, X.S.
Gene isolation and expression of a new Zn-finger
Unpublished
2 (bases 1 to 513)
Li, J.R., Wang, F., Li, Q.Z. and Zhang, X.S.
Direct Submission
Submitted (18-JUN-2002) College of Life Science, Shandong
Agricultural University, Dai zong street 61, Tai'an, Shandong
271018, P.R. China

FEATURES
Source
Location/Qualifiers
1..513
/organism="Triticum aestivum"
/mol_type="rRNA"
/cultiivar="PH 82-2-2"
/db_xref="taxon:4565"
/dev_stage="seeds"
/note="similar to HOF of Hordeum vulgare"

CDS
1..513
/note="similar to HOF of Hordeum vulgare"
/product="HOF"
/protein_id="AAP80610.1"
/db_xref="GI:32400756"
/translation="WVGKNGQELLEYPDYATKHAHAYPGHGMVILPSSAVG
/translating="WVGKNGQELLEYPDYATKHAHAYPGHGMVILPSSAVG
LMEERLHAKHIDORTDRTWQRRVPLPGSRPOLXGLAREKDEMTNRNCKSKR
LKYEMRSHNMVVAOMKMSQEDNOQLNLYKNKVKTEQSKVVEITLAVITQKLEETM
ENIFVRSKA"

ORIGIN

Query Match 10.2%; Score 192; DB 8; Length 513;
Best Local Similarity 62.3%; Pred. No. 8.5e-36;
Matches 319; Conservative 0; Mismatches 190; Indels 3; Gaps 1;
DB 1 TGGTTCGCGATGGGACCAAGAGCTGGAATCTTCAAGAGTTGAGGCTTGA 1023
964 TGGTTCGCGATGGGACCAAGAGCTGGAATCTTCAAGAGTTGAGGCTTGA 1023
1 TGGTTCGCGATGGGACCAAGAGCTTCTGATTTTCAAGAGTTGAGGCTTGA 60
DB 1024 GCAGCCATCTTCAATGTCACAGAGGCGATGGGATGATGTTCTGATTTGAGG 1083
61 GCAGCCATCTTCAATGTCACAGAGGCGATGGGATGATGTTCTGATTTGAGG 120
DB 1084 AGTGCATGCTATTTGAGGCGGACGCTTCAAGAGTTGAGGCTTGA 1143
121 TCTGCTGCTGCTATTTGAGGCGGACGCTTCAAGAGTTGAGGCTTGA 180
DB 1144 GATGATTTGCTG--GGGTCAAGAGGCGATGTTCTGAGGCTTGGCACTG 1200
181 GATGATTTGCTG--GGGTCAAGAGGCGATGTTCTGAGGCTTGGCACTG 240
DB 1201 TATGCTGCTGCTGCAAGAGGATGTCATATTCATCAACTCTTCAAGGCA 1260
241 TATGCTGCTGCTGCAAGAGGATGTCATATTCATCAACTCTTCAAGGCA 300
DB 1261 ACAAGGCTGAATTCGATGTAATCAATCAAGAGATGTTGTAAGAGGCTGAGG 1320
301 AGCGGCTGAATTCGATGTAATCAATCAAGAGATGTTGTAAGAGGCTGAGG 360
DB 1321 ATCTCTGAGACATCGAGCTGATCTTCAAGAGGCTTCAAGAGGCTGAGG 1380
361 ATCTCTGAGACATCGAGCTGATCTTCAAGAGGCTTCAAGAGGCTGAGG 420
DB 1381 CACGCGAAGTGTCTGAGGATCTCTGGAATTTGAGGAGAGGCTGAGGCTGCA 1440
421 CACGCGAAGTGTCTGAGGATCTCTGGAATTTGAGGAGAGGCTGAGGCTGCA 480
DB 1441 GAGGATATCGATCTGAGGAGAGGAGCTTAA 1472
481 GAGGATATCGATCTGAGGAGAGGAGCTTAA 512

RESULT 14

CNS08CAO/c 91053 bp DNA linear PLN 21-NOV-2003
LOCUS
DEFINITION
Oryza sativa chromosome 12, BAC OJ1561.A05 of library Monsanto
from chromosome 12 of cultivar Nipponbare of esp. Japonica of Oryza
sativa (rice), complete sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
HTG.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Eriocaridaceae; Oryzaceae; Oryza.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
Unpublished
2 (bases 1 to 91053)
Genoscope.
Direct Submission
Submitted (20-NOV-2003) Genoscope - Centre National de Séquençage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
On Mar 9, 2003 this sequence version replaced gi:28460590.
Center: Genoscope / Centre National de Séquençage

COMMENT

The following sequence is oriented from the T7 to the SP6 end. The
nucleotide sequence of this BAC clone was generated by combining
Monsanto, Syngenta and Genoscope sequencing data.
Upstream BAC (overlapping the T7 end) : OSJNB0029N15 (AC-BX35967)
Downstream BAC (overlapping the SP6 end) : OJ102_B11 (AC-AL713901)
----- Finishing boundaries -----
FINISHED SEGMENT STARTS AT BASE 1
FINISHED SEGMENT ENDS AT BASE 75085

FEATURES

Source
Location/Qualifiers
1..91053
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultiivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone_lib="Monsanto"

ORIGIN

Query Match 9.0%; Score 169.4; DB 8; Length 91053;
Best Local Similarity 61.1%; Pred. No. 2.6e-30;
Matches 310; Conservative 0; Mismatches 191; Indels 6; Gaps 2;
DB 460 TTTCCTGACGTGGAGATGATGATGATGATGATGATGATGATGATGATGAT 519
81521 TTTCCTGACGTGGAGATGATGATGATGATGATGATGATGATGATGATGAT 81462
DB 520 TTGGATGATTTCTGATGACGCTTCAAGATGATGATGATGATGATGATGAT 579
81461 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 81402
DB 580 AGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 639
81401 AGTTTGAAGCTCGGAAAATTCACAGTTTCTTCAAGGTTTCTTCAAGGCTG 81342
DB 640 TTGTCATGACGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 699
81341 TTAAGTGAAGACGTTTAAATGAAACCAATGACCAATGACCAATGACCAAT 81282

Qy 700 GGACCTGGTGCATCGATTGATATA---CTGCACCCCTCTACTAGCTCATGCGAGACA 756
 Db 81281 GGACCTGGTGCATCGATTGATATA---CTGCACCCCTCTACTAGCTCATGCGAGACA 81222
 Qy 757 AAAGAGCTAGGCGAGTTAAGCTCCATAGAGATTGGCTGAAGTTTAAAGAAAGATCTA 816
 Db 81221 AAGGGTTCTATTAAGGTCAAGGCTCACAGAAATGGCTAGTATTATGAGAAAGAGCTA 81162
 Qy 817 CAGATGAGAGGCGCATCTGTCATTCCTTGGTGAGATTATAGGCGAGGAGGCTTG 876
 Db 81161 TCTCCAGAGGGAATCTTCACTGATGATCTGGCGAACTTTAGGAAATGAAAGGATTTG 81102
 Qy 877 GGTGAGATGAAGAGATTATGAAATTTCTGGCTCAATGGTCAATCATCATGATACT 936
 Db 81101 CG---GAAAGACACGATCGTGAAGATTGTTGGCCACCAATGGTGTGTTATGAACACT 81045
 Qy 937 AGACTGATAGAGAGATTAAGATAG 963
 Db 81044 GTATTGAAACAAGATGAGATGATAG 81018

RESULT 15

CNS07Y01/c

LOCUS 112231 bp DNA linear PLN 21-NOV-2003

DEFINITION Oryza sativa chromosome 12, . BAC OJ1102_B11 of library Monsanto
 from chromosome 12 of cultivar Nipponbare of ssp. japonica of Oryza

sativa (rice), complete sequence.

ACCESSION AL713901.3 GI:38490465

VERSION HTG.

SOURCE Oryza sativa (japonica cultivar-group)
 Oryza sativa (japonica cultivar-group)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.REFERENCE 1 (bases 1 to 112231)
 Choinsse, N., Orjeda, G., Catolico, L., Demange, N., Wincker, P.,
 Segurens, B., Pelletier, F., Scarpelli, C., Salanoubat, M.,
 Weissenbach, J. and Quetier, F.

AUTHORS

TITLE Oryza sativa chromosome 12 sequencing
 Unpublished
 2 (bases 1 to 112231)

JOURNAL

REFERENCE

TITLES

JOURNAL

COMMENT

Direct Submission
 Submitted (20-NOV-2003) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : secretef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)On Nov 21, 2003 this sequence version replaced gi:23094326.
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: Secretef@genoscope.cns.fr

The following sequence is oriented from the T7 to the SP6 end. The
 nucleotide sequence of this BAC clone was generated by combining
 Monsanto and Genoscope sequencing data.
 Upstream BAC (overlapping the T7 end) : OJ1561_A05 (AC-AL831811)
 Downstream BAC (overlapping the SP6 end) : OJ1003_A04 (AC-AL713947)
 ----- Finishing boundaries
 FINISHED SEGMENT STARTS AT BASE 1
 FINISHED SEGMENT ENDS AT BASE 112231

FEATURES

source

Location/Qualifiers
 1..112231
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /sub_species="japonica"
 /db_xref="taxon:39947"
 /chromosome="12"
 /clone="OJ1102_B11"
 /clone_1fb="Monsanto"

ORIGIN

Query Match 9.0%; Score 169.4; DB 8; Length 112231;
 Best Local Similarity 61.1%; Pred. No. 2.6e-30;
 Matches 310; Conservative 0; Mismatches 191; Indels 6; Gaps 2;

Qy 460 TTTCCTGACGTGAGAGATGATGATTAATGCTTCTGAGAAAGAGATGATTCGATGCT 519
 Db 8800 TTGCTGCTGTTGATCTCGAACAAGATGATGATGATGATGATGATGATGATGAT 8741
 Qy 520 TTGATGATTTGATGAGACGCTTCGACAGTATGATGATGATGATGATGATGATGAT 579
 Db 8740 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8681
 Qy 580 ACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 639
 Db 8680 AGTTTGAAGCTCGGAAAAATCAAGTTTCTTGAAGTCTTGAAGTCTTGAAGTCTTGA 8621
 Qy 640 TTGTCATGACGACAGATTAATGATGATGATGATGATGATGATGATGATGATGATG 699
 Db 8620 TTAAGTGAAGACAGTAAATGATGATGATGATGATGATGATGATGATGATGATGAT 8561
 Qy 700 GGACCTGGTGCATCGATTGATATA---CTGCACCCCTCTACTAGCTCATGCGAGACA 756
 Db 8560 GGACCTGGTGCATCGATTGATATA---CTGCACCCCTCTACTAGCTCATGCGAGACA 8501
 Qy 757 AAAGAGCTAGGCGAGTTAAGCTCCATAGAGATTGGCTGAAGTTTAAAGAAAGATCTA 816
 Db 8500 AAGGGTTCTATTAAGGTCAAGGCTCACAGAAATGGCTAGTATTATGAGAAAGAGCTA 8441
 Qy 817 CAGATGAGAGGCGCATCTGTCATTCCTTGGTGAGATTATAGGCGAGGAGGCTTG 876
 Db 8440 TCTCCAGAGGGAATCTTCACTGATGATCTGGCGAACTTTAGGAAATGAAAGGATTTG 8381
 Qy 877 GGTGAGATGAAGAGATTATGAAATTTCTGGCTCAATGGTCAATCATCATGATACT 936
 Db 8380 CG---GAAAGACACGATCGTGAAGATTGTTGGCCACCAATGGTGTGTTATGAACACT 8324
 Qy 937 AGACTGATAGAGATTAAGATAG 963
 Db 8323 GTATTGAAACAAGATGAGATGATAG 8297

Search completed: March 29, 2004, 23:22:40
 Job time : 7093.65 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 08:05:52 ; Search time 4261.12 Seconds
(without alignments)
13161.120 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878

Sequence: 1 atgagttctagggtctgtctc.....ttcacatgaagatgatga 1878

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_hcc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hcc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	565	30.1	693	28	AQ959658 LERX20TF
2	534	28.4	606	28	AQ959657 LERX20TF
3	513	27.3	619	9	AV566465 AV566465
4	512	27.3	647	9	AU236368 AU236368

C	5	474	25.2	613	28	AQ964580	AQ964580 LERX20TF
	6	329	17.5	380	9	AV525508	AV525508 AV525508
	7	295	15.7	457	29	CNS00P9H	AL084227 Arabidops
	8	273	14.5	523	28	AQ959659	AQ959659 LERX20TF
	9	244	13.0	421	9	AU227277	AU227277 AU227277
	10	228	12.1	290	28	AQ010650	AQ010650 P27C8TRC
	11	169	9.0	548	9	AI999551	AI999551 701556368
	12	153	8.1	197	28	BH169457	BH169457 SALK_0013
	13	148	7.9	443	28	BH618393	BH618393 SALK_0390
	14	57	3.0	482	28	BH169448	BH169448 SALK_0013
	15	47	2.5	232	28	AQ964581	AQ964581 LERX20TF
	16	38	2.0	384	29	BK653272	BK653272 Arabidops
	17	38	2.0	437	29	BK287410	BK287410 Arabidops
	18	29	1.5	644	14	CD835509	CD835509 BM45_0451
	19	29	1.5	697	28	BH535855	BH535855 BOHR13TF
	20	28	1.5	587	13	BO507998	BO507998 EST615413
	21	24	1.3	123	29	BK654916	BK654916 Arabidops
	22	24	1.3	405	29	BK287411	BK287411 Arabidops
	23	22	1.2	270	13	BY023513	BY023513 BY023513
	24	22	1.2	562	12	BM262455	BM262455 daq41b10
	25	22	1.2	621	12	BJ625937	BJ625937 BJ625937
	26	22	1.2	675	14	CA798850	CA798850 EST1648 M
	27	22	1.2	708	14	CD815300	CD815300 BN15_0250
	28	22	1.2	896	29	CG969741	CG969741 MBHR11TR
	29	22	1.2	967	9	AV254322	AV254322 AV254322
	30	22	1.2	2684	11	AK029504	AK029504 Mus muscu
	31	22	1.1	258	10	BB578249	BB578249 BB578249
	32	21	1.1	319	12	BG279367	BG279367 b2f06np.f
	33	21	1.1	350	14	CF598325	CF598325 NCEST3489
	34	21	1.1	376	12	BG543396	BG543396 E0910 Chi
	35	21	1.1	396	14	CB831627	CB831627 NCEST3463
	36	21	1.1	402	10	BF248666	BF248666 NCEST3401
	37	21	1.1	409	14	CF371330	CF371330 NCEST3450
	38	21	1.1	414	10	BF824285	BF824285 NCEST3434
	39	21	1.1	419	10	BF249331	BF249331 NCEST3409
	40	21	1.1	424	10	BF716900	BF716900 NCEST3418
	41	21	1.1	425	14	CF260387	CF260387 NCEST3404
	42	21	1.1	432	10	BF716619	BF716619 NCEST3415
	43	21	1.1	432	10	BF249140	BF249140 NCEST3407
	44	21	1.1	433	12	BG235108	BG235108 NCEST3440
	45	21	1.1	434	10	BF248638	BF248638 NCEST3401

ALIGNMENTS

RESULT 1
AQ959658 693 bp DNA linear GSS 28-JAN-2000
LOCUS LERX20TFB LERX Arabidopsis thaliana genomic clone LERX20TF, genomic
DEFINITION survey sequence.

ACCESSION AQ959658
VERSION AQ959658.1 GI:6787359

KEYWORDS GSS.
SOURCE Arabidopsis thaliana (chale cress)

ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

TITLE
Buell, C.R., Lin, X., Pat, G., Barnstead, M., Bowman, C., Ullrich, T.,
Feildhym, T., Liang, F., Greasy, T. and Fraser, C.M.
Genomic survey sequencing of Landsberg erecta ecotype of
Arabidopsis thaliana and identification of sequence-based
polymorphisms

JOURNAL
COMMENT unpublished (2000)

Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: atc@tigr.org
For additional information, see <http://www.tigr.org/tcdb/at/at.html>

Seq primer: TF
Class: shotgun.
Location/Qualifiers

1. 693
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="LANDBERG ERECTA"
/db_xref="taxon:3702"
/clone="LBERM21"
/note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.6-0.8 Kbp before ligation."

ORIGIN

Query Match 30.1%; Score 565; DB 28; Length 693;
Best Local Similarity 99.8%; Pred. No. 1.6e-276;
Matches 615; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 134 AGGTCAATTTCCAGAGAAACAAGAAACAACCACTTCTGGAAAACTTGGGTTT 193
DB 688 AGGTCAATTTCCAGAGAAACAAGAAACAACCACTTCTGGAAAACTTGGGTTT 629
QY 194 CTCAGAAATTCGAATCCTCTAGAGCTTGGGGTGTGTCAGAGCAAGGAGGAGGAGCAAG 253
DB 628 CTCAGAAATTCGAATCCTCTAGAGCTTGGGGTGTGTCAGAGCAAGGAGGAGGAGCAAG 569
QY 254 TATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGTCGGGCAATTCAG 313
DB 568 TATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATGTCGGGCAATTCAG 509
QY 314 CTAACTATATCTGTCGGGGAGCAAGCTTGAAGCAAGAAAGTATGATTAACAATTGGGCA 373
DB 508 CTAACTATATCTGTCGGGGAGCAAGCTTGAAGCAAGAAAGTATGATTAACAATTGGGCA 449
QY 374 CCCCACTGATATCTGCGCCCTCTTGAAGAGATGAAATGGCAGCAAGAGAGTT 433
DB 448 CCCCACTGATATCTGCGCCCTCTTGAAGAGATGAAATGGCAGCAAGAGAGTT 389
QY 434 CTGCTCAGACACACAGCTGTGTCAGAGAGTTCTCTGACGTGAGAGTATGTGATATGCTT 493
DB 388 CTGCTCAGACACACAGCTGTGTCAGAGAGTTCTCTGACGTGAGAGTATGTGATATGCTT 329
QY 494 CTGAGAGAGAGATGATTCGATGCTTGGATATTCGATGACGACCTTGAAGTANG 553
DB 328 CTGAGAGAGAGATGATTCGATGCTTGGATATTCGATGACGACCTTGAAGTANG 269
QY 554 ATTATGCTGAGATGTGAGTCAAAAGAGCCATGATCAAGAAAGCAAAATTAAGTGTCA 613
DB 268 ATTATGCTGAGATGTGAGTCAAAAGAGCCATGATCAAGAAAGCAAAATTAAGTGTCA 209
QY 614 AAAAGTCTTTGGAGCTTGGATAGCTTGTGATCGAGCAGATTAATGAACAAGAGC 673
DB 208 AAAAGTCTTTGGAGCTTGGATAGCTTGTGATCGAGCAGATTAATGAACAAGAGC 149
QY 674 AGTGGCAATTTCCAGCTTGTCAAGACGAGCTGTCATCGATTAACCTTGACC 733
DB 148 AGTGGCAATTTCCAGCTTGTCAAGACGAGCTGTCATCGATTAACCTTGACC 89
QY 734 CTCTACTAGCTCATGC 749
DB 88 CTCTACTAGCTCATGC 73

```

RESULT 2
AQ959657/c 606 bp DNA linear GSS 28-JAN-2000
LOCUS
DEFINITION LBERM21TF LBER Arabidopsis thaliana genomic clone LBERM21, genomic survey sequence.
ACCESSION AQ959657
VERSION AQ959657.1 GI:6787358
KEYWORDS GSS.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished (2000)
Contact: Xiaoying Lin
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0200
Email: atc@ligr.org
For additional information, see <http://www.tigr.org/tdb/at/ac.html>
Seq primer: TF

Class: shotgun.
Location/Qualifiers

1. 606
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="LANDBERG ERECTA"
/db_xref="taxon:3702"
/clone="LBERM21"
/note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was sheared to 0.6-0.8 Kbp before ligation."

ORIGIN

Query Match 28.4%; Score 534; DB 28; Length 606;
Best Local Similarity 100.0%; Pred. No. 1e-260;
Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 216 AGCTTGGGGTGTGTCAGACAGAGAGAGTACCACTATCTGGAGAGAGAAACATGT 275
DB 606 AGCTTGGGGTGTGTCAGACAGAGAGAGTACCACTATCTGGAGAGAGAAACATGT 547
QY 276 ATCCGAGAGAGTAAACGGCAATGTCGGGGCACTTCAACTAATATCTGGCGGGAG 335
DB 546 ATCCGAGAGAGTAAACGGCAATGTCGGGGCACTTCAACTAATATCTGGCGGGAG 487
QY 336 AGCGTTGACAGAAATGTATTAACAATTGGGCAACCCCACTGTATCTGCGCTCC 395
DB 486 AGCGTTGACAGAAATGTATTAACAATTGGGCAACCCCACTGTATCTGCGCTCC 427
QY 486 AGCGTTGACAGAAATGTATTAACAATTGGGCAACCCCACTGTATCTGCGCTCC 455
DB 396 TTTGAGAGAGAGTAAATTTGGCAGGCAAGAGAGTCTGCTCAGACACAGCTGTGCA 427
QY 426 TTTGAGAGAGAGTAAATTTGGCAGGCAAGAGAGTCTGCTCAGACACAGCTGTGCA 367
QY 456 GGAATTTCTGACGTGAGAGATGATGTGATTAATGCTCTGAGGAAGAAATGATTCGA 515
DB 366 GGAATTTCTGACGTGAGAGATGATGTGATTAATGCTCTGAGGAAGAAATGATTCGA 307
QY 516 TGCTTTGATGATTCGATGACGACCTTGCAAGTATTAATGATCTCGATGTGATCA 575
DB 306 TGCTTTGATGATTCGATGACGACCTTGCAAGTATTAATGATCTCGATGTGATCA 247
QY 576 AAAAGCCATGATCAAGAAAGCAATTAAGTGTTCAAAAGTCTTTGGCAGCTTGA 635
DB 246 AAAAGCCATGATCAAGAAAGCAATTAAGTGTTCAAAAGTCTTTGGCAGCTTGA 187
QY 636 TAGCTTGTGATCGACCAATTAATGAACACAGAGGCGATGCGCATTCAGCTTGTCA 695
DB 186 TAGCTTGTGATCGACCAATTAATGAACACAGAGGCGATGCGCATTCAGCTTGTCA 127
QY 696 GAAAGCACTGTGCGCATGATGATTAATGATCACTGCACTCTTAAGTCAATGC 749
DB 126 GAAAGCACTGTGCGCATGATGATTAATGATCACTGCACTCTTAAGTCAATGC 73

```

RESULT 3

AV566465/c 619 bp mRNA linear EST 07-SEP-2000
 LOCUS AV566465
 DEFINITION Arabidopsis thaliana green siliques Columbia Arabidopsis
 thaliana cDNA clone SQ244b06F 3', mRNA sequence.
 ACCESSION AV566465
 VERSION AV566465.1 GI:8737917
 SOURCE EST.
 ORGANISM Arabidopsis thaliana (thale cress)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 619)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 DNA Res. 7 (3), 175-180 (2000)
 JOURNAL MEDLINE PUBMED
 COMMENT 10907847
 Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
 Location/Qualifiers
 1..619
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="SQ244b06F"
 /tissue_type="green siliques"
 /clone_1ib="Arabidopsis thaliana green siliques Columbia"
 /note="Vector: pBluescriptII SK-, Site_1: EcoRI; Site_2:
 XhoI"

ORIGIN
 Query Match 27.3%; Score 513; DB 9; Length 619;
 Best Local Similarity 100.0%; Pred. No. 5.5e-250;
 Matches 513; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1291 CAAGAGATGTTTAAAGAGAGAGAGATCTTGAGACATCAGACGCTGAACTAC 1350
 |||||
 DB 619 CAAGAGATGTTTAAAGAGAGAGAGATCTTGAGACATCAGACGCTGAACTAC 560
 |||||
 QY 1351 TTTTAAAGAACGCTCTCAAAACAGAACAGACGCCAAGGTGTTGAGGAATCTCTGAA 1410
 |||||
 DB 559 TTTTAAAGAACAGCTCTCAAAACAGAACAGACGCCAAGGTGTTGAGGAATCTCTGAA 500
 |||||
 QY 1411 ATTATGAGGAG 1470
 |||||
 DB 499 ATTATGAGGAG 440
 |||||
 QY 1471 AAGATGCAAGATGAAACAGAAACAGGAAAGATGATGACACAGACAGTTTTCATGAT 1530
 |||||
 DB 439 AAGATGCAAGATGAAACAGAAACAGGAAAGATGATGACACAGACAGTTTTCATGAT 380
 |||||
 QY 1531 TCAATCAAAACAGATCCATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGATGTTGAG 1590
 |||||
 DB 379 TCAATCAAAACAGATCCATGAAAGAGAGAGAGAGAGAGAGAGAGAGATGTTGAG 320
 |||||
 QY 1591 CAGCAGGAACGTGCAAGGTGTTGGCCAGCAGCAGCAGCAATTAATCCCTCTGACAT 1650
 |||||
 DB 319 CAGCAGGAACGTGCAAGGTGTTGGCCAGCAGCAGCAGCAATTAATCCCTCTGACAT 260
 |||||
 QY 1651 GACGATGCGCAAG 1710
 |||||
 DB 259 GACGATGCGCAAG 200
 |||||
 QY 1711 ATGAGAGAGTTTGTGAG 1770
 |||||
 DB 199 ATGAGAGAGTTTGTGAG 140
 |||||

QY 1771 GACATGAAGAGAGAGATCAGAGAGATATTT 1803
 |||||
 DB 139 GACATGAAGAGAGATCAGAGAGATATTT 107
 |||||

RESULT 4
 LOCUS AU236368
 DEFINITION AU236368 647 bp mRNA linear EST 01-APR-2002
 thaliana cDNA clone RAFL14-93-K05.5',
 mRNA sequence.
 ACCESSION AU236368
 VERSION AU236368.1 GI:19875537
 SOURCE EST.
 ORGANISM Arabidopsis thaliana (thale cress)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 647)
 Seki, M., Narusaka, M., Ishida, J., Kamita, A., Saitou, M., Nakajima, M.,
 Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
 Arakawa, T., Shibata, K., Shinagawa, A., Muramatsu, M., Hayashizaki, Y.
 and Shinozaki, K.
 Large scale analysis of Arabidopsis full-length cDNA (2002b)
 Unpublished (2002)
 Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060
 Email: mseki@rtc.riken.go.jp
 An Arabidopsis full-length cDNA library was constructed essentially
 as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
 and XhoI was ligated to modified lambda PhiC-1 vector (Carninci et
 al., submitted for publication) digested with BamHI and SalI. This
 clone is in a modified pBluescript vector. Please visit our web
 site (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
 details.

ORIGIN
 Query Match 27.3%; Score 512; DB 9; Length 647;
 Best Local Similarity 100.0%; Pred. No. 1.8e-249;
 Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGAGTTTCAAGGCTGCTCCCAATGCTTAAGAAAGAAAGCTTCAGGCTGTTAAGCCCT 60
 |||||
 DB 92 AAGAGTTTCAAGGCTGCTCCCAATGCTTAAGAAAGAAAGCTTCAGGCTGTTAAGCCCT 151
 |||||
 QY 61 GAGTTTGAACAGTTGTTTCAAGTTTGGCAGGAGAGAGAGAGAGAGAGAGAGAGAGAT 120
 |||||
 DB 152 GAGTTTGAACAGTTGTTTCAAGTTTGGCAGGAGAGAGAGAGAGAGAGAGAGAGAT 211
 |||||
 QY 121 GGAGAGAGTGGAGAGCTATTTCAGAGAGAAACAGAACACACAGAGAAACCTCTGGA 180
 |||||
 DB 212 GGAGAGAGTGGAGAGCTATTTCAGAGAGAAACAGAACACACAGAGAAACCTCTGGA 271
 |||||
 QY 181 AAAAATTGCTTCTCAAGATTCGAATCTCTTAAGCTTGGGCTGTCAGAGCAAGGG 240
 |||||
 DB 272 AAAAATTGCTTCTCAAGATTCGAATCTCTTAAGCTTGGGCTGTCAGAGCAAGGG 331
 |||||
 QY 241 AGAGTTAGCAAGTCTCTGAGAGAGAAACAATGTATCCGGAGAGATTAACGGCAATGGT 300
 |||||

Db 332 AGAGTACCAAGTATCTGGAGAGAGAAACATGATATCGGAGAGGTAAAGGCAATGCT 391
 Qy 301 CGGGGATTCACAGTAAATATCTGGTGGGAGAGAGGTTGAGAGAGAAATGATATAC 360
 Db 332 CGGGGATTCACAGTAAATATCTGGTGGGAGAGAGGTTGAGAGAGAAATGATATAC 451
 Qy 361 AACTTTGTGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGATGAAATTGGCAG 420
 Db 452 AACTTTGTGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGATGAAATTGGCAG 511
 Qy 421 GCAAGAGAGGTTCTGCTTACGACACAGCTGTGAGAGGTTTCCGAGTGGAGATGAT 480
 Db 512 GCAAGAGAGGTTCTGCTTACGACACAGCTGTGAGAGGTTTCCGAGTGGAGATGAT 571
 Qy 481 GTGATTAATGCTTCTGAGAGAGAGATGATTC 512
 Db 572 GTGATTAATGCTTCTGAGAGAGAGATGATTC 603

RESULT 5
 A0964580/c 613 bp DNA linear GSS 28-JAN-2000
 LOCUS LERGX20TF LERG Arabidopsis thaliana genomic clone LERGX20, genomic
 DEFINITION survey sequence.
 ACCESSION A0964580
 VERSION A0964580.1 GI:6792281
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 613)
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Ulterbach, T.,
 Feldhym, T., Liang, F., Creasy, T. and Fraser, C.M.
 TITLE Genomic survey sequencing of landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms

JOURNAL
 COMMENT Unpublished (2000)
 Contact: Xiaoying Lin
 The Institute for Genomic Research
 912 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atc@tigr.org
 For additional information, see <http://www.tigr.org/cdb/ac/at.html>
 Seq primer: TF
 Class: shotgun.

FEATURES

Source
 1..613
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERECTA"
 /db_xref="taxon:3702"
 /clone="LERGX20"
 /clone_lib="LERGX"
 /note="Organ: Leaf; Vector: pUC19K; Total genomic DNA was
 sheared to 0.4-0.7 Kbp before ligation."

ORIGIN

Query Match 25.2%; Score 474; DB 28; Length 613;
 Best Local Similarity 99.8%; Pred. No. 4.5e-230;
 Matches 524; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAGAAAGAGCTTCAAGGTTGTTAAGGCTT 60
 Db 525 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAGAAAGAGCTTCAAGGTTGTTAAGGCTT 466
 Qy 61 GAGGTTGAACAGTTGTTCAAGGTTTGGCAGGAGAGAGAGCTTCTTCAAGATGAT 120
 Db 465 GAGGTTGAACAGTTGTTCAAGGTTTGGCAGGAGAGAGAGCTTCTTCAAGATGAT 406

Qy 121 GGAGAGAGTGGAGGATCTTTCCAGAAAGAACAAACAGAAACACTTCTGCA 180
 Db 405 GGAGAGAGTGGAGGATCTTTCCAGAAAGAACAAACAGAAACACTTCTGCA 346
 Qy 181 AAAAATTGGGTTTCTCAGAAATTCGAATCTCTTAAGACTTTGGGTTGTGACAGCAAGG 240
 Db 345 AAAAATTGGGTTTCTCAGAAATTCGAATCTCTTAAGACTTTGGGTTGTGACAGCAAGG 286
 Qy 241 AGAGGTGCAACGATATCTGGAGAGAGAGAAACATGTATCCGGAGAGGTAAAGGCAATGCT 300
 Db 285 AGAGGTGCAACGATATCTGGAGAGAGAGAAACATGTATCCGGAGAGGTAAAGGCAATGCT 226
 Qy 301 CGGGGATTCACAGTAAATATCTGTGCGGAGAGAGCTTTGAGAGAGAGATGAAATTGGCAG 360
 Db 225 CGGGGATTCACAGTAAATATCTGTGCGGAGAGAGCTTTGAGAGAGATGAAATTGGCAG 166
 Qy 361 AACTTTGTGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGAGATGAAATTGGCAG 420
 Db 165 AACTTTGTGGACACCCCACTGTATCTGCCCTCTTTGGAAGAGAGATGAAATTGGCAG 106
 Qy 421 GCAAGAGAGGTTCTGCTTACGACACAGCTGTGAGAGGTTTCTCTGACGTGAGAGATGAT 480
 Db 105 GCAAGAGAGGTTCTGCTTACGACACAGCTGTGAGAGGTTTCTCTGACGTGAGAGATGAT 46
 Qy 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGAT 525
 Db 45 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGAT 1

RESULT 6
 AV525508 380 bp mRNA linear EST 01-SEP-2000
 LOCUS AV525508 Arabidopsis thaliana aboveground organs two to six-week
 DEFINITION old Arabidopsis thaliana cDNA clone APD5d02R 5', mRNA sequence.
 ACCESSION AV525508.1 GI:8685036
 VERSION AV525508
 KEYWORDS EST.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana

REFERENCE
 AUTHORS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 380)
 Asamizu, E., Nakamura, Y., Sato, S. and Tabata, S.
 TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation
 of 12,028 non-redundant expressed sequence tags from normalized and
 size-selected cDNA libraries
 JOURNAL DNA Res. 7 (3), 175-180 (2000)
 MEDLINE 20363093
 PUBMED 10907847

COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute
 Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: <http://www.kazusa.or.jp/en/plant/>.

FEATURES

Source
 1..380
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="APD5d02R"
 /tissue_type="aboveground organs"
 /dev_stage="two to six-week old"
 /clone_lib="Arabidopsis thaliana aboveground organs two to
 six-week old"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
 XhoI"

ORIGIN

Query Match 17.5%; Score 329; DB 9; Length 380;
 Best Local Similarity 99.7%; Pred. No. 4e-156;
 Matches 379; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 276 ATCCGGAGAGTAACGCGAATGTCGGGCAATTCAGCTAACATATCTGATCGGGAGC 335
 DB 1 ATCCGGAGAGTAACGCGAATGTCGGGCAATTCAGCTAACATATCTGATCGGGAGC 60
 QY 336 ACCGTTGACAGAAAGTATGATPAACAATCTTTGGGACCCCACTGATCTCGCCCTCC 395
 DB 61 ACCGTTGACAGAAAGTATGATPAACAATCTTTGGGACCCCACTGATCTCGCCCTCC 120
 QY 396 TTGGAGAGAGAGTGAATTTGGAGGAGAGAGAGGTTGCTGACGACACAGCTGTGCA 455
 DB 121 TTGGAGAGAGAGTGAATTTGGAGGAGAGAGAGGTTGCTGACGACACAGCTGTGCA 180
 QY 456 GGAGTTTCTGACGTCGAGAGATGATGTGATTAATGCTTCTGAGAGAGAGATGATTCGA 515
 DB 181 GGAGTTTCTGACGTCGAGAGATGATGTGATTAATGCTTCTGAGAGAGAGATGATTCGA 240
 QY 516 TCTCTTGATGATTTCTGATGACGACCTTGCAAGTATGATTAATGCTGAGATGTGATCA 575
 DB 241 TCTCTTGATGATTTCTGATGACGACCTTGCAAGTATGATTAATGCTGAGATGTGATCA 300
 QY 576 AAGAGCCATGATCAAGAGAGAGATTAAGTGTTCAGAAAGTCTTTGGAGCTTGA 635
 DB 301 AAGAGCCATGATCAAGAGAGAGATTAAGTGTTCAGAAAGTCTTTGGAGCTTGA 360
 QY 636 TAGCTTGTCATCGACGACA 655
 DB 361 TAGCTTGTCATCGACGACA 380

RESULT 7
 LOCUS CNS00P9H 457 bp DNA linear GSS 28-JUN-1999
 DEFINITION Arabidopsis thaliana genome survey sequence T7 end of BAC F821. of
 IGF library from strain Columbia of Arabidopsis thaliana, genomic
 survey sequence.
 AL084227
 AL084227.1 GI:5285367
 GSS.
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 457)
 Salanoubat,M., Choisme,N., Artiguenave,F., Brottier,P., Wincker,P.,
 Samson,D., Saurin,W., Weissenbach,J. and Quetier,F.
 Unpublished
 2 (bases 1 to 457)
 Genoscope.
 Direct Submission
 Submitted (25-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Location/Qualifiers
 1..457
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="F8G21"
 /clone_1lb="IGF"
 /note="end : T7"

ORIGIN
 Query Match 15.7%; Score 295; DB 29; Length 457;
 Best Local Similarity 100.0%; Pred. No. 9.8e-139;
 Matches 295; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 AGTGCTCGGAGTGGCAACGAGCTGCTGGAATCTTCAACAAGTATGAGGCTTTA 1021
 DB 55 AGTGCTCGGAGTGGCAACGAGCTGCTGGAATCTTCAACAAGTATGAGGCTTTA 114

QY 1022 GAGCAGCCATTCCTATGCTCCAGAGGCCATCTGGAGTAGTGTTCATGATTTGACA 1081
 DB 115 GAGCAGCCATTCCTATGCTCCAGAGGCCATCTGGAGTAGTGTTCATGATTTGACA 174
 QY 1082 GAGTGCCACTGGCTATTTTGAAGGCCGAAACGCTTCCACCGGAGTTAGCTGAGATGGGT 1141
 DB 175 GAGTGCCACTGGCTATTTTGAAGGCCGAAACGCTTCCACCGGAGTTAGCTGAGATGGGT 234
 QY 1142 TAGATGAATTCCTGGGGTGAAGAGCGAGATGTTTCTGAGAGGTGTTCGCCAATCT 1201
 DB 235 TAGATGAATTCCTGGGGTGAAGAGCGAGATGTTTCTGAGAGGTGTTCGCCAATCT 294
 QY 1202 ATGCTTCTTCCGCAAGAGCAAGATCTGACATATTCATCAACTCTCAAG 1256
 DB 295 ATGCTTCTTCCGCAAGAGCAAGATCTGACATATTCATCAACTCTCAAG 349

RESULT 8
 LOCUS A0959659 523 bp DNA linear GSS 28-JAN-2000
 DEFINITION LEREM21TR LERE Arabidopsis thaliana genomic clone LEREM21, genomic
 survey sequence.
 A0959659
 A0959659.1 GI:6787360
 GSS.
 Arabidopsis thaliana (thale cress)
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 523)
 Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
 Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
 Genomic survey sequencing of landberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms
 Unpublished (2000)
 Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atc@igf.org
 For additional information, see <http://www.tigr.org/cdb/at/at.html>
 Seq primer: TR
 Claes: Shotgun.

ORIGIN
 Location/Qualifiers
 1..523
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERECTA"
 /db_xref="taxon:3702"
 /clone="LEREM21"
 /clone_1lb="LERE"
 /note="Organ: Leaf; Vector: pUC19UK; Total genomic DNA was
 sheared to 0.6-0.8 Kbp before ligation."

Query Match 14.5%; Score 273; DB 28; Length 523;
 Best Local Similarity 99.3%; Pred. No. 1.8e-127;
 Matches 423; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 167 GAAACACTTCTGGAAAACTTGGGTTTCTCGAATTCGAATCTCTAGAGCTTGGGGTG 226
 DB 25 GAAACACTTCTGGAAAACTTGGGTTTCTCGAATTCGAATCTCTAGAGCTTGGGGTG 84
 QY 227 GTGACGACGAGAGAGAGTAGCAAGTATCTGGAGAGAGAAACATATATCCGGAGAG 286
 DB 85 GTGACGACGAGAGAGAGTAGCAAGTATCTGGAGAGAGAAACATATATCCGGAGAG 144
 QY 287 GTAACGCAATGCTGCGGGCAATTCAGGTTAATATCTGCTGCGGAGAGAGAGCTTGACA 346
 DB 145 GTAACGCAATGCTGCGGGCAATTCAGGTTAATATCTGCTGCGGAGAGAGAGCTTGACA 204

QY 347 GAAAGTATGATACAACTTTGTGCAACCCCACTGATCTGCGCCCTCTTTGGAAGAG 406
 DB 205 GAAAGTATGATACAACTTTGTGCAACCCCACTGATCTGCGCCCTCTTTGGAAGAG 264
 QY 407 GATGGAATTTGGAGGAGGAGAGGTTCTGCTCAGCAACAGCTGTGAGAGATTTCCCG 466
 DB 265 GATGGAATTTGGAGGAGGAGAGGTTCTGCTCAGCAACAGATGTGAGAGATTTCCCG 324
 QY 467 ACCTGAGAGATGATGATGATTAATCTCTTCTGAGGAAGAGATTCGATCTTTGGATG 526
 DB 325 ACCTGAGAGATGATGATGATTAATCTCTTCTGAGGAAGAGATTCGATCTTTGGATG 384
 QY 527 ATTCTGATGACGACCTTGGAGATGATGATTAATCTCTGATGTGATGATCAAAAGGCGCATG 586
 DB 385 ATTCTGATGACGACCTTGGAGATGATGATTAATCTCTGATGTGATGATCAAAAGGCGCATG 444
 QY 587 GATCAC 592
 DB 445 GATCAC 450

RESULT 9
 AU227277/c 421 bp mRNA linear EST 23-APR-2002
 LOCUS AU227277 RAF14 Arabidopsis thaliana cDNA clone RAF14-93-K05 3'
 DEFINITION mRNA sequence.

ACCESSION AU227277 GI:19741924
 VERSION AU227277
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 421)
 Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Saitou, M., Nakajima, M.,
 Akiyama, K., Enju, A., Oono, Y., Sakurai, T., Carninci, P., Kawai, Y.,
 Itoh, M., Ishii, Y., Arakawa, T., Shibata, K., Shinagawa, A.,
 Muramatsu, M., Hayashizaki, Y. and Shinozaki, K.

JOURNAL Large scale analysis of Arabidopsis full-length cDNA
 Unpublished (2002)
 COMMENT Contact: Motoaki Seki
 Plant Functional Genomics Research Group
 RIKEN Genomic Sciences Center
 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-4359
 Fax: 81-298-36-9060

FEATURES
 source
 1. 421
 /organism="Arabidopsis thaliana"
 /mol_type="mRNA"
 /db_xref="taxon:3702"
 /clone="RAF14-93-K05"
 /issue_type="root"
 /lab_host="DH10B"
 /clone_lib="RAF14"
 /note="Site_1: BamHI; Site_2: SalI"

ORIGIN

Query Match 13.0%; Score 244; DB 9; Length 421;
 Best local similarity 100.0%; Pred. No. 1e-112;
 Matches 244; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1635 TAATCCCTTAGCATGACGATTCGCAAGAGAGAGCTGAGAGAGTGTCAAGCTTCATGCA 1694

DB 421 TAATCCCTTAGCATGACGATTCGCAAGAGAGAGCTGAGAGAGTGTCAAGCTTCATGCA 362
 QY 1695 GTTTCAGAGAGAGAGATGAGAGATTTGTGTAAGAGAGAGATGCTATTAAGATCA 1754
 DB 361 GTTTCAGAGAGAGAGATGAGAGATTTGTGTAAGAGAGAGATGCTATTAAGATCA 302
 QY 1755 AGAGAGAGATGAGAGAGATGAGAGAGATTTGTGTAAGAGAGATTTGTGATCTGAGAG 1814
 DB 301 AGAGAGAGATGAGAGAGATGAGAGAGATTTGTGTAAGAGAGATTTGTGATCTGAGAG 242
 QY 1815 AGAATTTGATGAGCTTTGGAACAGCTCATGTAACAGATGGCTTCAAAATGAGATGA 1874
 DB 241 AGAATTTGATGAGCTTTGGAACAGCTCATGTAACAGATGGCTTCAAAATGAGATGA 182
 QY 1875 TTGA 1878
 DB 181 TTGA 178

RESULT 10
 AQ010650 290 bp DNA linear GSS 29-MAY-1998
 LOCUS F27C8TRC IGF Arabidopsis thaliana genomic clone F27C8, genomic
 DEFINITION survey sequence.

ACCESSION AQ010650 GI:3165927
 VERSION AQ010650
 KEYWORDS Arabidopsis thaliana (thale cress)
 SOURCE Arabidopsis thaliana
 ORGANISM Arabidopsis thaliana
 Arabidopsis thaliana; Streptophyta; Embryophyta; Tracheophyta;
 Eukaryota; Viridiplantae; Magnoliophyta; eudicotyledons; core eudicots;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 290)
 Rounsley, S.D., Suh, E.J., Wible, C., Golden, K., Shatsman, S., Choi, P.,
 Yu, K., Aktineloye, B., Shen, K., Goonasekaram, S., Millicher, J.,
 Adams, M.D. and Venter, J.C.

JOURNAL A BAC End Sequence Database for Identifying Minimal Overlaps in
 Arabidopsis Genomic Sequencing. Update 4
 Unpublished (1998)
 COMMENT Other GSSs: F27C8TRC
 Contact: Steve Rounsley
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: rounsley@igr.org

Seg primer: M13 Reverse
 Class: BAC ends
 High quality sequence stop: 290.

FEATURES
 source
 1. 290
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /clone="F27C8"
 /sex="hermaphrodite"
 /clone_lib="IGF"
 /note="Vector: BelobACII; Site_1: EcoRI; Site_2: EcoRI;
 Produced by Thomas Altmann"

ORIGIN

Query Match 12.1%; Score 228; DB 28; Length 290;
 Best local similarity 100.0%; Pred. No. 1.3e-104; Indels 0; Gaps 0;
 Matches 228; Conservative 0; Mismatches 0;

QY 962 AGTGCTCTGCAATGGGACCAAGAGCTGCTGGAATCTTCAACAAGATGAGGCTCTTA 1021
 DB 63 AGTGCTCTGCAATGGGACCAAGAGCTGCTGGAATCTTCAACAAGATGAGGCTCTTA 122

QY 1022 GAGCAGCGCATTCCTATGCTGCACAGGCGCATGCTGGATGAGATGCTTTGAGA 1081

Db 123 GAGCAGCCATTCCTATGCTCCACAGGGGCATCGTGGATGAGTGTTCGATGTTTGACA 182

Qy 1082 GGAGTGGCCACTGCTGATTTGAGAGCGCAGCCTCCACCGGGAGTTAGCTGAGATGGGCT 1141

Db 183 GCGAGTGGCCACTGCTGATTTGAGAGCGCAGCCTCCACCGGGAGTTAGCTGAGATGGGCT 242

Qy 1142 TAGATAGATTCCTGCGGGTCAGAACCGCAGTATGTTTCTGAGGCTG 1189

Db 243 TAGATAGATTCCTGCGGGTCAGAACCGCAGTATGTTTCTGAGGCTG 290

RESULT 11 548 bp mRNA linear EST 08-SEP-1999

LOCUS A1999551/1 701556368 A. thaliana, Columbia Col-0, rosette-3 Arabidopsis

DEFINITION thaliana cDNA clone 701556368, mRNA sequence.

ACCESSION A1999551

VERSION A1999551.1 GI:5846456

KEYWORDS EST.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 548)

AUTHORS Chen, J., Momiyama, M., Chan, B., Mooney, M., Carroun, B., Gilliland, D., Wang, X., Hillman, J., Guegler, K., Kim, C., Doyle, M., Bracke, P., Gorgone, G., Burns, D., Griffin, J., Mouanoutoua, M., Nguyen, D., Tan, R., Rose, M., Warren, B., Ton, B., Kastury, K., Bottino, C., Cardpio, T., Policky, J., Suzuki, G., Argentine, C., Shah, S., Noblidge, A., Murry, L., Turner, C., Krikorian, S., Elder, L., and Hanson, D.

TITLE Arabidopsis thaliana Gene Expression Microarray

JOURNAL Unpublished (1999)

COMMENT Contact: David Smolter, Ph.D. Genome Systems, Inc., a wholly owned subsidiary of Incyte Pharmaceuticals, Inc. 4633 World Parkway Circle, St. Louis, MO 63134, USA Tel: 877-577-2733 Fax: 314-427-3324 Email: service@genomesystems.com.

FEATURES

source

1..548

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/cultivar="Columbia Col-0"

/db_xref="taxon:3702"

/clone="701556368"

/tissue_type="rosette"

/dev_stage="4 - 7 weeks"

/clone_lib="A. thaliana, Columbia Col-0, rosette-3"

/note="Vector: pSPORT Site_1: NotI; Site_2: SalI; cDNA library was derived from untreated rosette tissue from Arabidopsis thaliana, Columbia Col-0, at 4 - 7 weeks. Plants were grown in 1:1:1 peat moss/vermiculite/perlite soil at 22 deg. C +/- 3 deg. C under constant light, and watered with fertilizer. cDNA synthesis was initiated using a NotI-oligo(dT) primer. Double-stranded cDNA was blunted, ligated to SalI adaptors, digested with NotI, size-selected, and cloned into the NotI and SalI sites of the pSPORT vector."

ORIGIN

Query Match 9.0%; Score 169; DB 9; Length 548;

Best Local Similarity 100.0%; Pred. No. 2.2e-74;

Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1720 GATGAGAGGTTTGGAGAGAGAGAGTGTCTAATAAAGATCAAGAGAGAGATGGA 1769

Db 345 GATGAGAGGTTTGGAGAGAGAGAGTGTCTAATAAAGATCAAGAGAGAGATGGA 286

Qy 1770 AGACATGAAGAAGAGCATCAAGAGAGATATTGATCTGAGAAAGATTGATGAGGC 1829

Db 285 AGACATGAAGAAGAGCATCAAGAGAGATATTGATCTGAGAAAGATTGATGAGGC 226

Qy 1830 TTGGAAACGCTCATGTACAAGCATGGCCTTCAATGAAGATGATTGA 1878

Db 225 TTGGAAACGCTCATGTACAAGCATGGCCTTCAATGAAGATGATTGA 177

RESULT 12 197 bp DNA linear GSS 03-OCT-2001

LOCUS BH169457

DEFINITION SALK_001394 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_001394, genomic survey sequence.

ACCESSION BH169457

VERSION BH169457.1 GI:15904832

KEYWORDS GSS.

SOURCE Arabidopsis thaliana (thale cress)

ORGANISM Arabidopsis thaliana

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 197)

AUTHORS Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J., and Ecker, J.R.

TITLE A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salik.edu

This is single pass sequence recovered from the left border of TDNA.

FEATURES

source

1..197

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/strain="Columbia 0"

/db_xref="taxon:3702"

/clone="SALK_001394"

/clone_lib="Arabidopsis thaliana TDNA insertion lines"

/note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html

ORIGIN

Query Match 8.1%; Score 153; DB 28; Length 197;

Best Local Similarity 100.0%; Pred. No. 2.3e-66;

Matches 153; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 50 GTTATAGGCTGAGCTGTAACAGTTGTTCAAGTTTGGCAGGACGAGACTGGCTTCT 109

Db 1 GTTATAGGCTGAGCTGTAACAGTTGTTCAAGTTTGGCAGGACGAGACTGGCTTCT 60

Qy 110 CACAAGATGATGAGAGAGAGTGGAGTCAATTCACAAGAACAGAACAAACAGAGA 169

Db 61 CACAAGATGATGAGAGAGAGTGGAGTCAATTCACAAGAACAGAACAAACAGAGA 120

Qy 170 ACACTTGTGAAAAAATTGGGTTTCTCAGAAAT 202

Db 121 ACACTTGTGAAAAAATTGGGTTTCTCAGAAAT 153

RESULT 13

BH169457/1

LOCUS BH618393 443 bp DNA linear GSS 30-JAN-2002
 DEFINITION SALK_039005 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_039005, genomic survey sequence.
 ACCESSION BH618393
 VERSION BH618393.1 GI:18428488
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 443)
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrihab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shin, P., Zimmerman, J., and Ecker, J.R.
 Shin, P., Zimmerman, J., and Ecker, J.R.
 A Sequence-indexed library of Insertion Mutations in the Arabidopsis Genome
 Unpublished (2001)
 JOURNAL Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of TDNA. This sequence lies within an annotated intron of At5g23570.
 Class: TDNA tagged.
 Location/Qualifiers
 1..443
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_039005"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
 Query Match 7.9%; Score 148; DB 28; Length 443;
 Best Local Similarity 100.0%; Pred. No. 1.1e-63;
 Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 962 AGTGGCTGGCATGGCAACCAAGAGCTGCGAATGATGACAAGATGAGGCTCTTA 1021
 |||||
 DB 407 AGTGGCTGGCATGGCAACCAAGAGCTGCGAATGATGACAAGATGAGGCTCTTA 348
 |||||
 QY 1022 GACACAGCCATCTCTATGTCACACAGGCGCATGTCGAGTGTCTGATGTTTGA 1081
 |||||
 DB 347 GAGACGCCATCTCTATGTCACACAGGCGCATGTCGAGTGTCTGATGTTTGA 288
 |||||
 QY 1082 GCAGTGCCTACTGCTATTGAGAGCGCA 1109
 |||||
 DB 287 GCAGTGCCTACTGCTATTGAGAGCGCA 260
 |||||

RESULT 14
 BH169448 482 bp DNA linear GSS 03-OCT-2001
 LOCUS SALK_001377 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_001377, genomic survey sequence.
 DEFINITION BH169448
 ACCESSION BH169448.1 GI:15904823
 VERSION BH169448
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE
 1 (bases 1 to 482)
 Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrihab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shin, P., Zimmerman, J., and Ecker, J.R.
 Shin, P., Zimmerman, J., and Ecker, J.R.
 A Sequence-indexed library of Insertion Mutations in the Arabidopsis Genome
 Unpublished (2001)
 JOURNAL Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of TDNA.
 Class: TDNA tagged.
 Location/Qualifiers
 1..482
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_001377"
 /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"

ORIGIN
 Query Match 3.0%; Score 57; DB 28; Length 482;
 Best Local Similarity 100.0%; Pred. No. 3.2e-17;
 Matches 57; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 AGAAGACACAGACCAACCAAGAGAGCTGCGAATGATGACAAGATGAGGCTCTTA 202
 |||||
 DB 111 AGAAGACACAGACCAACCAAGAGAGCTGCGAATGATGACAAGATGAGGCTCTTA 167
 |||||

RESULT 15
 AQ964581 232 bp DNA linear GSS 28-JAN-2000
 LOCUS LERGX20TR LERX Arabidopsis thaliana genomic clone LERGX20, genomic survey sequence.
 DEFINITION AQ964581
 ACCESSION AQ964581.1 GI:6792282
 VERSION AQ964581
 KEYWORDS GSS.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 232)
 Buell, C.R., Lin, X., Pai, G., Barnstead, M., Bowman, C., Uterbach, T., Feldblum, T., Liang, F., Cressy, T., and Fraser, C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of Arabidopsis thaliana and identification of sequence-based polymorphisms
 Unpublished (2000)
 JOURNAL Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atc@igr.org
 For additional information, see <http://www.tigr.org/cdb/at/at.html>
 Seq primer: TR
 Class: shotgun.
 Location/Qualifiers

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 08:22:12 ; Search time 140.312 Seconds
(without alignments)
7427.693 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878
Sequence: 1 atgagctctaggctgctgc.....ttcacatgaagatgatga 1878

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents, NA:*
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/ECTUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	19	1.0	481	4 US-09-615-192A-109	Sequence 109, App
C 2	19	1.0	481	4 US-09-169-789-109	Sequence 109, App
C 3	19	1.0	495	2 US-08-975-316-27	Sequence 27, Appl
C 4	19	1.0	495	4 US-09-615-192A-27	Sequence 27, Appl
C 5	19	1.0	495	4 US-09-169-789-27	Sequence 27, Appl
6	19	1.0	2000	1 US-09-041-075A-4	Sequence 4, Appl
7	19	1.0	2000	1 US-09-041-075A-6	Sequence 6, Appl
8	19	1.0	8257	4 US-09-595-684B-30	Sequence 30, Appl
9	19	1.0	8503	4 US-09-620-312D-130	Sequence 130, Appl
10	10	1.0	1664976	4 US-08-916-421B-1	Sequence 1, Appl
11	18	1.0	462	4 US-09-489-039A-5315	Sequence 5315, Ap
C 12	18	1.0	1254	4 US-09-322-478-25	Sequence 25, Appl
C 13	18	1.0	1829	1 US-07-688-352C-15	Sequence 15, Appl
C 14	18	1.0	1829	3 US-08-474-379C-15	Sequence 15, Appl
C 15	18	1.0	1829	3 US-09-146-249A-15	Sequence 15, Appl
C 16	18	1.0	1829	3 US-08-206-188B-15	Sequence 15, Appl
C 17	18	1.0	1829	5 PCT-US91-02714-15	Sequence 15, Appl
C 18	18	1.0	1952	4 US-09-533-028-89	Sequence 89, Appl
C 19	18	1.0	2001	4 US-09-489-039A-5217	Sequence 5217, Ap
C 20	18	1.0	3031	1 US-08-785-241-2	Sequence 2, Appl
C 21	18	1.0	7478	4 US-10-104-966-15	Sequence 15, Appl
C 22	18	1.0	8906	2 US-08-826-267-1	Sequence 1, Appl
C 23	18	1.0	35524	3 US-08-923-137-1	Sequence 1, Appl
C 24	18	1.0	42571	4 US-09-810-347-3	Sequence 3, Appl
C 25	18	1.0	100848	4 US-09-596-002-39	Sequence 39, Appl
C 26	17	0.9	20	4 US-09-851-896-62	Sequence 62, Appl
C 27	17	0.9	47	4 US-09-422-978-3322	Sequence 3322, Ap

28	17	0.9	247	4 US-09-411-977-5	Sequence 5, Appl
C 29	17	0.9	396	4 US-09-107-532A-767	Sequence 767, App
C 30	17	0.9	428	4 US-09-702-705-1214	Sequence 1214, Ap
C 31	17	0.9	428	4 US-09-736-457-1214	Sequence 1214, Ap
C 32	17	0.9	428	4 US-09-614-124B-1214	Sequence 1214, Ap
C 33	17	0.9	428	4 US-09-671-325-1214	Sequence 1214, Ap
C 34	17	0.9	477	4 US-09-621-976-10972	Sequence 10972, A
C 35	17	0.9	536	4 US-09-621-976-16220	Sequence 16220, A
C 36	17	0.9	561	1 US-08-565-386-3	Sequence 3, Appl
C 37	17	0.9	576	4 US-09-107-532A-1728	Sequence 1728, Ap
C 38	17	0.9	793	4 US-09-221-017B-940	Sequence 940, App
C 39	17	0.9	1250	3 US-09-018-584A-36	Sequence 36, Appl
C 40	17	0.9	1284	4 US-09-328-352-594	Sequence 594, Appl
C 41	17	0.9	1290	4 US-09-543-681A-1061	Sequence 1061, Ap
C 42	17	0.9	1323	4 US-09-489-039A-4961	Sequence 4961, Ap
C 43	17	0.9	1401	4 US-09-134-000C-2355	Sequence 2355, Ap
C 44	17	0.9	1428	2 US-08-960-022-7	Sequence 7, Appl
45	17	0.9	1496	4 US-09-016-434-1127	Sequence 1127, Ap

ALIGNMENTS

```

RESULT 1
US-09-615-192A-109/c
; Sequence 109, Application US/09615192A
; Patent No. 6410718
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c4U
; CURRENT APPLICATION NUMBER: US/09/615,192A
; CURRENT FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; NUMBER OF SEQ ID NOS: 405
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-09-615-192A-109

Query Match          1.0% Score 19; DB 4; Length 481;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      74 TGGTCAAGGTTTGCAGG 92
Db      470 TGGTCAAGGTTTGCAGG 452

RESULT 2
US-09-169-789-109/c
; Sequence 109, Application US/09169789
; Patent No. 6653528
; GENERAL INFORMATION:
; APPLICANT: Bloksberg, Leonard N.
; APPLICANT: Havukkala, Ilkka
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003c2
; CURRENT APPLICATION NUMBER: US/09/169,789
; CURRENT FILING DATE: 1998-10-09
; EARLIER APPLICATION NUMBER: US 08/975,316
; EARLIER FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: US 08/713,000

```

EARLIER FILING DATE: 1996-09-11
NUMBER OF SEQ ID NOS: 185
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 109
LENGTH: 481
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-169-789-109

Query Match 1.0%; Score 19; DB 4; Length 481;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92
DB 470 TGGTCAAGGTTGGCAGG 452

RESULT 3

US-08-975-316-27/c
Sequence 27, Application US/08975316
Patent No. 5952486
GENERAL INFORMATION:
APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
APPLICANT: and GRIERSON, Alastair W.
TITLE OF INVENTION: MATERIALS AND METHODS FOR
THE MODIFICATION OF PLANT LIGNIN CONTENT
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,316
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/713,000
FILING DATE: September 11, 1996
ATTORNEY/AGENT INFORMATION:
NAME: SLEATH, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000/1003C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-975-316-27

Query Match 1.0%; Score 19; DB 2; Length 495;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92
DB 484 TGGTCAAGGTTGGCAGG 466

RESULT 4

US-09-615-192A-27/c
Sequence 27, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
Modification of Plant Lignin Content
FILE REFERENCE: 11000.1003C4U
CURRENT APPLICATION NUMBER: US/09/615,192A
CURRENT FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/713,000
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 27
LENGTH: 495
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-615-192A-27

Query Match 1.0%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92
DB 484 TGGTCAAGGTTGGCAGG 466

RESULT 5

US-09-169-789-27/c
Sequence 27, Application US/09169789
Patent No. 6653528
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
APPLICANT: Havukkala, Ilkka
TITLE OF INVENTION: Materials and Methods for the
Modification of Plant Lignin Content
FILE REFERENCE: 11000.1003C2
CURRENT APPLICATION NUMBER: US/09/169,789
CURRENT FILING DATE: 1998-10-09
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1997-11-21
EARLIER APPLICATION NUMBER: US 08/713,000
EARLIER FILING DATE: 1996-09-11
NUMBER OF SEQ ID NOS: 185
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 27
LENGTH: 495
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-169-789-27

Query Match 1.0%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGGTTGGCAGG 92
DB 484 TGGTCAAGGTTGGCAGG 466

RESULT 6

US-09-041-075A-4
Sequence 4, Application US/09041075A
Patent No. H002022
GENERAL INFORMATION:
APPLICANT: Heidler, Steven A

```

; APPLICANT: Radding, Jeffrey A
; TITLE OF INVENTION: IPC SYNTBASE GENES FROM FUNGI
; FILE REFERENCE: X-11242 Sequence Lat
; Patent No. H002022
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Candida krusei
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (300)..(1739)
US-09-041-075A-4
```

```

Query Match          1.0%; Score 19; DB 1; Length 2000;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      523 GATGATTCTGATGACGACC 541
      |||||||
DB      1560 GATGATTCTGATGACGACC 1578
```

```

RESULT 7
US-09-041-075A-6
; Sequence 6, Application US/09041075A
; Patent No. H002022
; GENERAL INFORMATION:
; APPLICANT: Heidler, Steven A
; APPLICANT: Radding, Jeffrey A
; TITLE OF INVENTION: IPC SYNTBASE GENES FROM FUNGI
; FILE REFERENCE: X-11242 Sequence Lat
; Patent No. H002022
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2000
; TYPE: RNA
; ORGANISM: Candida krusei
US-09-041-075A-6
```

```

Query Match          1.0%; Score 19; DB 1; Length 2000;
Best Local Similarity 73.7%; Pred. No. 12;
Matches 14; Conservative 5; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      523 GATGATTCTGATGACGACC 541
      |||||||
DB      1560 GAUGAUUCUGAUGACGACC 1578
```

```

RESULT 8
US-09-595-684B-30
; Sequence 30, Application US/09595684B
; Patent No. 6544766
; GENERAL INFORMATION:
; APPLICANT: Berard, Christophe
; APPLICANT: Ohashi, Cara
; APPLICANT: Sakowicz, Roman
; APPLICANT: Vaisberg, Eugeni
; APPLICANT: Wood, Kenneth
; APPLICANT: Yu, Ming
; TITLE OF INVENTION: Human kinesins and methods of producing
; TITLE OF INVENTION: and purifying human kinesins
; FILE REFERENCE: cytop036
```

```

; CURRENT APPLICATION NUMBER: US/09/595,684B
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 8257
; TYPE: DNA
; ORGANISM: Human
US-09-595-684B-30
```

```

Query Match          1.0%; Score 19; DB 4; Length 8257;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1802 TTGATCTGGAGAAAGATT 1820
      |||||||
DB      2141 TTGATCTGGAGAAAGATT 2159
```

```

RESULT 9
US-09-620-312D-130
; Sequence 130, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Abundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; TITLE OF INVENTION: polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pc_fl_genes Version 1.0
; SEQ ID NO 130
; LENGTH: 8503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (91)..(8082)
US-09-620-312D-130
```

```

Query Match          1.0%; Score 19; DB 4; Length 8503;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1802 TTGATCTGGAGAAAGATT 1820
      |||||||
DB      2141 TTGATCTGGAGAAAGATT 2159
```

```
RESULT 10
```

US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Built et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g

```
/ NAME/KEY: misc feature
/ LOCATION: (1310988) ..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1313224) ..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349473) ..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1349491) ..(1349491)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1470091) ..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1569020) ..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1602912) ..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1603734) ..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1637998) ..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
/ NAME/KEY: misc feature
/ LOCATION: (1664854) ..(1664854)
/ OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1
```

```
Query Match 1.0%; Score 19; DB 4; Length 1664976;
Best Local Similarity 100.0%; Pred.No. 17;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1702 GAGAAAGATGAGAGACT 1720
DB 358045 GAGAAAGATGAGAGACT 358063
```

```
RESULT 11
US-09-489-039A-5315
/ Sequence 5315, Application US/09489039A
/ Patent No. 6610836
/ GENERAL INFORMATION:
/ APPLICANT: Gary Breton et. al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/ FILE REFERENCE: 2709 2004001
/ CURRENT FILING DATE: 2000-01-27
/ PRIOR APPLICATION NUMBER: US 60/117,747
/ PRIORITY FILING DATE: 1999-01-29
/ NUMBER OF SEQ ID NOS: 14342
/ SEQ ID NO 5315
/ LENGTH: 462
/ TYPE: DNA
/ ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5315
```

```
Query Match 1.0%; Score 18; DB 4; Length 462;
Best Local Similarity 100.0%; Pred.No. 37;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1616 GCCAGCAGCAGCAACA 1633
DB 216 GCCAGCAGCAGCAACA 233
```

```
RESULT 12
US-09-322-478-25/C
/ Sequence 25, Application US/09322478
```

```
/ Patent No. 6331662
/ GENERAL INFORMATION:
/ APPLICANT: Wright, David A.
/ APPLICANT: Voytas, Daniel F.
/ TITLE OF INVENTION: Plant Retroelements and Methods Related Thereto
/ FILE REFERENCE: P-1065 ISUP Plant Retroelement
/ CURRENT FILING DATE: 1999-05-28
/ EARLIER APPLICATION NUMBER: US/09/322,478
/ EARLIER FILING DATE: 1998-05-29
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 25
/ LENGTH: 1254
/ TYPE: DNA
/ ORGANISM: Pisum sativum
US-09-322-478-25
```

```
Query Match 1.0%; Score 18; DB 4; Length 1254;
Best Local Similarity 100.0%; Pred.No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1335 TCAGCAGCTGAACACTT 1352
DB 890 TCAGCAGCTGAACACTT 873
```

```
RESULT 13
US-07-688-352C-15
/ Sequence 15, Application US/07688352C
/ Patent No. 5527896
/ GENERAL INFORMATION:
/ APPLICANT: Wigler, Michael H.
/ APPLICANT: Colicelli, John J.
/ TITLE OF INVENTION: Cloning By Complementation and Related
/ NUMBER OF SEQUENCES: 57
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Marshall, O'Toole, Gerstein, Murray &
/ STREET: Two First National Plaza, 20 South Clark
/ CITY: Chicago
/ STATE: Illinois
/ COUNTRY: USA
/ ZIP: 60603
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/07/688,352C
/ FILING DATE: 19910419
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 07/511,715
/ FILING DATE: 20-APR-1990
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Borun, Michael F.
/ REGISTRATION NUMBER: 25447
/ REFERENCE/DOCKET NUMBER: 27805/30197
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (312) 346-5750
/ TELEFAX: (312) 984-9740
/ TELEX: 25-3856
/ INFORMATION FOR SEQ ID NO: 15:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1829 base pairs
/ TYPE: NUCLEIC ACID
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
```

FEATURE:
NAME/KEY: CDS
LOCATION: 30..1421
US-07-688-352C-15

Query Match 1.0%; Score 18; DB 1; Length 1829;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 ATTCCAGAGAGACAG 156
DB 985 ATTCCAGAGAGACAG 1002

RESULT 14
US-08-474-379C-15
Sequence 15, Application US/08474379C
Patent No. 5977305
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: CLONING BY COMPLEMENTATION AND RELATED
PROCESSES
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,379C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/206,188
FILING DATE: 01-MAR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/588,352
FILING DATE: 19-APR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
REFERENCE/DOCKET NUMBER: 27866/32771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1829 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1421
US-08-474-379C-15

Query Match 1.0%; Score 18; DB 2; Length 1829;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 ATTCCAGAGAGACAG 156

DB 985 ATTCCAGAGAGACAG 1002

RESULT 15
US-09-146-249A-15
Sequence 15, Application US/09146249A
Patent No. 6069240
GENERAL INFORMATION:
APPLICANT: Wigler, Michael H.
APPLICANT: Colicelli, John J.
TITLE OF INVENTION: Cloning by Complementation and Related
Processes
NUMBER OF SEQUENCES: 85
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/146,249A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/511,715
FILING DATE: 20-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: Clough, David W.
REGISTRATION NUMBER: 36,107
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1829 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 30..1421
US-09-146-249A-15

Query Match 1.0%; Score 18; DB 3; Length 1829;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 139 ATTCCAGAGAGACAG 156
DB 985 ATTCCAGAGAGACAG 1002

Search completed: March 30, 2004, 18:11:52
Job time : 149.312 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 30, 2004, 14:48:39 ; Search time 1169.15 Seconds
(without alignments)
5980.258 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878
Sequence: 1 atgagctcagggctgctcc.....ttcacacatgaagatgatga 1878

Scoring table: OLIGO NUC
Gapop 60.0 , Gapext 60.0

Searched: 2458946 seqs, 1861504846 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCOT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCOT_NEW_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	1.1	1377	US-10-424-599-91177	Sequence 91177, A
2	20	1.1	1589	US-10-424-599-91176	Sequence 91176, A
3	19	1.0	172	US-10-085-783A-19246	Sequence 19246, A
4	19	1.0	172	US-10-242-535A-19246	Sequence 19246, A
5	19	1.0	397	US-10-424-599-57837	Sequence 57837, A
6	19	1.0	429	US-10-027-632-293476	Sequence 293476, A
7	19	1.0	481	US-10-174-693-109	Sequence 109, App
8	19	1.0	495	US-10-174-693-27	Sequence 27, App
9	19	1.0	501	US-09-833-790-167	Sequence 167, App
10	19	1.0	681	US-10-424-599-18549	Sequence 18549, A
11	19	1.0	696	US-10-027-632-23555	Sequence 23555, A
12	19	1.0	696	US-10-027-632-23556	Sequence 23556, A
13	19	1.0	824	US-10-027-632-11811	Sequence 11811, A
14	19	1.0	824	US-10-027-632-11812	Sequence 11812, A
15	19	1.0	1137	US-10-369-493-24736	Sequence 24736, A

16	19	1.0	1341	9	US-09-938-842A-2465	Sequence 2465, Ap
17	19	1.0	1341	11	US-09-938-842A-2465	Sequence 2465, Ap
18	19	1.0	1576	12	US-10-424-599-4367	Sequence 4367, Ap
19	19	1.0	2000	9	US-09-742-582-4	Sequence 4, Appli
20	19	1.0	2000	9	US-09-742-582-6	Sequence 6, Appli
21	19	1.0	2000	10	US-09-742-580-4	Sequence 4, Appli
22	19	1.0	2000	10	US-09-742-580-6	Sequence 6, Appli
23	19	1.0	2000	10	US-09-742-581-4	Sequence 4, Appli
24	19	1.0	2000	10	US-09-742-581-6	Sequence 6, Appli
25	19	1.0	2240	15	US-10-108-260A-1451	Sequence 1451, Ap
26	19	1.0	3579	14	US-10-176-847-9	Sequence 9, Appli
27	19	1.0	6930	10	US-09-770-107-1	Sequence 1, Appli
28	19	1.0	7992	10	US-09-893-519A-140	Sequence 140, App
29	19	1.0	8491	14	US-10-133-013-260	Sequence 260, App
30	19	1.0	8493	13	US-10-071-766-51	Sequence 51, Appli
31	19	1.0	8503	14	US-10-037-270-130	Sequence 130, App
32	19	1.0	8503	15	US-10-117-722-130	Sequence 130, App
33	19	1.0	47448	15	US-10-085-117-145	Sequence 145, App
34	19	1.0	1691139	14	US-10-067-514-1	Sequence 1, Appli
35	19	1.0	1691139	15	US-10-419-723-1	Sequence 1, Appli
36	19	1.0	2940917	15	US-10-027-632-174763	Sequence 174763, A
37	18	1.0	182	14	US-10-029-386-26108	Sequence 26108, A
38	18	1.0	189	9	US-09-864-761-32008	Sequence 32008, A
39	18	1.0	209	9	US-09-960-352-12756	Sequence 12756, A
40	18	1.0	256	12	US-10-424-599-116017	Sequence 116017, A
41	18	1.0	270	9	US-09-960-352-14374	Sequence 14374, A
42	18	1.0	304	12	US-10-424-599-112786	Sequence 112786, A
43	18	1.0	312	9	US-09-728-445-2930	Sequence 290, App
44	18	1.0	315	9	US-09-960-352-12961	Sequence 12961, A
45	18	1.0	395	9	US-09-960-352-3954	Sequence 3954, Ap

ALIGNMENTS

RESULT 1
US-10-424-599-91177
; Sequence 91177, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 91177
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1
US-10-424-599-91177

Query Match 1.1%; Score 20; DB 12; Length 1377;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 673 CAGTGGCATTTGTCAGCTTG 692
DB 359 CAGTGGCATTTGTCAGCTTG 378

RESULT 2
US-10-424-599-91176
; Sequence 91176, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J

```
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 91176
LENGTH: 1589
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
US-10-424-599-91176
```

```
Query Match          1.0%; Score 19; DB 12; Length 1589;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      673 CAGTGGCATTGTCACGCTTG 692
      |||
Db      359 CAGTGGCATTGTCACGCTTG 378
```

```
RESULT 3
US-10-085-783A-19246
Sequence 19246, Application US/10085783A
Publication No. US20040037841A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2002
CURRENT APPLICATION NUMBER: US/10/085,783A
PRIOR FILING DATE: 2002-02-28
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 19246
LENGTH: 172
TYPE: DNA
ORGANISM: Human
US-10-085-783A-19246
```

```
Query Match          1.0%; Score 19; DB 12; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1767 GGAAGACATGAGAGAGG 1785
      |||
Db      86 GGAAGACATGAGAGAGG 104
```

```
RESULT 4
US-10-242-535A-19246
Sequence 19246, Application US/10242535A
Publication No. US20040013663A1
GENERAL INFORMATION:
APPLICANT: ChondroGene Inc.
APPLICANT: Ilew, C.C.
TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
FILE REFERENCE: 4231/2005
CURRENT APPLICATION NUMBER: US/10/242,535A
CURRENT FILING DATE: 2002-09-12
PRIOR APPLICATION NUMBER: US 10/085,783
PRIOR FILING DATE: 2002-02-28
```

```
PRIOR APPLICATION NUMBER: US 60/305,340
PRIOR FILING DATE: 2001-07-13
PRIOR APPLICATION NUMBER: US 60/275,017
PRIOR FILING DATE: 2001-03-12
PRIOR APPLICATION NUMBER: US 60/271,955
PRIOR FILING DATE: 2001-02-28
NUMBER OF SEQ ID NOS: 58994
SOFTWARE: PatentIn version 3.2
SEQ ID NO 19246
LENGTH: 172
TYPE: DNA
ORGANISM: Human
US-10-242-535A-19246
```

```
Query Match          1.0%; Score 19; DB 15; Length 172;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1767 GGAAGACATGAGAGAGG 1785
      |||
Db      86 GGAAGACATGAGAGAGG 104
```

```
RESULT 5
US-10-424-599-57837
Sequence 57837, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 57837
LENGTH: 397
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_23239C.1
US-10-424-599-57837
```

```
Query Match          1.0%; Score 19; DB 12; Length 397;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1700 AAGAGAAAGATGAGGA 1718
      |||
Db      262 AAGAGAAAGATGAGGA 280
```

```
RESULT 6
US-10-027-632-293476
Sequence 293476, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
```

```

; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; PRIOR APPLICATION NUMBER: US 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 293476
; LENGTH: 429
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-293476
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 15; Length 429;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 488 ATGCTTCTGAGAGAGAA 506
DB 76 ATGCTTCTGAGAGAGAA 94
```

```
RESULT 7
US-10-174-693-109/c
; Sequence 109, Application US/10174693
; Publication No. US2003013173A1
; GENERAL INFORMATION:
; APPLICANT: Blockberg, Leonard N.
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003C5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 109
; LENGTH: 481
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-174-693-109
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 14; Length 481;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 74 TGGTCAAGGTTGGCAG 92
DB 470 TGGTCAAGGTTGGCAG 452
```

```
RESULT 8
US-10-174-693-27/c
; Sequence 27, Application US/10174693
; Publication No. US2003013173A1
; GENERAL INFORMATION:
; APPLICANT: Blockberg, Leonard N.
; TITLE OF INVENTION: Materials and Methods for the
; FILE REFERENCE: 11000.1003C5
; CURRENT APPLICATION NUMBER: US/10/174,693
; CURRENT FILING DATE: 2002-06-18
```

```

; PRIOR APPLICATION NUMBER: US 08/975,316
; PRIOR FILING DATE: 1997-11-21
; PRIOR APPLICATION NUMBER: US 08/713,000
; PRIOR FILING DATE: 1996-09-11
; PRIOR APPLICATION NUMBER: US 09/169,789
; PRIOR FILING DATE: 1998-10-09
; PRIOR APPLICATION NUMBER: US 09/615,192
; PRIOR FILING DATE: 2000-07-12
; NUMBER OF SEQ ID NOS: 407
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 495
; TYPE: DNA
; ORGANISM: Eucalyptus grandis
US-10-174-693-27
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 14; Length 495;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 74 TGGTCAAGGTTGGCAG 92
DB 484 TGGTCAAGGTTGGCAG 466
```

```
RESULT 9
US-09-833-790-167
; Sequence 167, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tonglong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raedoh
; APPLICANT: Indrias, Carol Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 167
; LENGTH: 501
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-833-790-167
```

```
Query Match
Best Local Similarity 100.0%; Score 19; DB 9; Length 501;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1802 TTGATCTGAGAGAGATT 1820
DB 10 TTGATCTGAGAGAGATT 28
```

```
RESULT 10
US-10-424-599-18549/c
; Sequence 18549, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
```

```
SEQ ID NO 18549
LENGTH: 681
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_116754C.1
US-10-424-599-18549
```

```
Query Match      1.0%; Score 19; DB 12; Length 681;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1745 TAAAGATCAGAGAA 1763
DB      371 TAAAGATCAGAGAA 353
```

```
RESULT 11
US-10-027-632-2355/c
Sequence 23555, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
```

```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
```

```
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23555
LENGTH: 696
TYPE: DNA
ORGANISM: Human
US-10-027-632-23555
```

```
Query Match      1.0%; Score 19; DB 15; Length 696;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1122 GGAGTTAGCTGAGTGGG 1140
DB      82 GGAGTTAGCTGAGTGGG 64
```

```
RESULT 12
US-10-027-632-23556/c
Sequence 23556, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
```

```
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23556
LENGTH: 696
TYPE: DNA
ORGANISM: Human
US-10-027-632-23556
```

```
QY      1122 GGAGTTAGCTGAGTGGG 1140
DB      82 GGAGTTAGCTGAGTGGG 64
```

```
RESULT 13
US-10-027-632-11811
Sequence 11811, Application US/10027632
Publication No. US20030204075A9
GENERAL INFORMATION:
```

```
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
POLYMORPHISMS IN THE HUMAN GENOME
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11811
LENGTH: 824
TYPE: DNA
ORGANISM: Human
US-10-027-632-11811
```

```
Query Match      1.0%; Score 19; DB 15; Length 824;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1491 CAGGGAAGAGATGATGCA 1509
DB      579 CAGGGAAGAGATGATGCA 597
```

```
RESULT 14
US-10-027-632-11812
Sequence 11812, Application US/10027632
```

Job time : 1176.15 secs

```
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108627.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11812
; LENGTH: 824
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-11812
```

```
Query Match          1.0%; Score 19; DB 15; Length 824;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1491 CAGGAGAGATGATGCA 1509
          |||||
Db       579 CAGGAGAGATGATGCA 597
```

```
RESULT 15
US-10-369-493-24736
; Sequence 24736, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052) B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 24736
; LENGTH: 1137
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
US-10-369-493-24736
```

```
Query Match          1.0%; Score 19; DB 15; Length 1137;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1702 GAGAAAGATGAGAGAGT 1720
          |||||
Db       607 GAGAAAGATGAGAGAGT 625
```

Search completed: March 30, 2004, 23:20:33

This Page Blank (uspio)

CC SGR3 results in plants with increased resistance to viruses, while
 CC inactivation of SGR3 in transgenic plants (e.g. by expressing antisense
 CC RNA, by mutation or by homologous recombination) increases the level of
 CC the transgene product. This product may e.g. impart resistance (to
 CC herbicide, insects or pathogens), alter contents of essential fatty acids
 CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or
 CC interferon

SQ Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 100.0%; Score 1878; DB 4; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAAAAGAGCTTCAGAGGTGTTATAGGCT 60
 DB 1 ATGAGTTCTAGGGCTGTGCTCAATGCTTAAGAAAAGAGCTTCAGAGGTGTTATAGGCT 60
 QY 61 GAGGTTGAACAGTTGGTTCAAGTTTGGCAGGACGAGACTGGCTTTTCAAGATGAT 120
 DB 61 GAGGTTGAACAGTTGGTTCAAGTTTGGCAGGACGAGACTGGCTTTTCAAGATGAT 120
 QY 121 GAGAGAGAGTGGAGGTCATTTCAGAAAGAAACAAACAAACAAACAAACAACTTCTGA 180
 DB 121 GAGAGAGAGTGGAGGTCATTTCAGAAAGAAACAAACAAACAAACAAACAACTTCTGA 180
 QY 121 GAGAGAGAGTGGAGGTCATTTCAGAAAGAAACAAACAAACAAACAAACAACTTCTGA 180
 DB 181 AAAAATTGGGTTTCTCGAATTCGAATTCCTCTAGAGCTTGGGGTGTGACAGCAAGG 240
 QY 181 AAAAATTGGGTTTCTCGAATTCGAATTCCTCTAGAGCTTGGGGTGTGACAGCAAGG 240
 DB 181 AAAAATTGGGTTTCTCGAATTCGAATTCCTCTAGAGCTTGGGGTGTGACAGCAAGG 240
 QY 241 AGAGGTAGCAACGTATCTGAGAGAGAGAAACATGATCCGGAGAGAGTAAACGCAATG 300
 DB 241 AGAGGTAGCAACGTATCTGAGAGAGAGAAACATGATCCGGAGAGAGTAAACGCAATG 300
 QY 241 AGAGGTAGCAACGTATCTGAGAGAGAGAAACATGATCCGGAGAGAGTAAACGCAATG 300
 DB 301 CGGGGCAATTCAGCTTAACATATCTGCTCGGGAGAGAGCTTGAAGCAAAATGATTAAC 360
 QY 301 CGGGGCAATTCAGCTTAACATATCTGCTCGGGAGAGAGCTTGAAGCAAAATGATTAAC 360
 DB 301 CGGGGCAATTCAGCTTAACATATCTGCTCGGGAGAGAGCTTGAAGCAAAATGATTAAC 360
 QY 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTGGCAG 420
 DB 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTGGCAG 420
 QY 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTGGCAG 420
 DB 361 AACTTTGGGCAACCCCACTGATCTCGCCCTCTTTGAAAGAGATGAAATTGGCAG 420
 QY 421 GCAAGAGAGGTTTCTGCTCAGACACACAGCTGTGAGAGGTTTCTGACGTGAGAGAT 480
 DB 421 GCAAGAGAGGTTTCTGCTCAGACACACAGCTGTGAGAGGTTTCTGACGTGAGAGAT 480
 QY 421 GCAAGAGAGGTTTCTGCTCAGACACACAGCTGTGAGAGGTTTCTGACGTGAGAGAT 480
 DB 481 GTGATTAATGCTTCTGAGAGAGAGAAATGATCCGATGCTTTGATGATTCGATGAGAC 540
 QY 481 GTGATTAATGCTTCTGAGAGAGAGAAATGATCCGATGCTTTGATGATTCGATGAGAC 540
 DB 481 GTGATTAATGCTTCTGAGAGAGAGAAATGATCCGATGCTTTGATGATTCGATGAGAC 540
 QY 541 CTTCGAAAGTATGATTAATGATCTCGAGTGTGAGTCAAAAGAGCAATGATCGAAAGCAG 600
 DB 541 CTTCGAAAGTATGATTAATGATCTCGAGTGTGAGTCAAAAGAGCAATGATCGAAAGCAG 600
 QY 541 CTTCGAAAGTATGATTAATGATCTCGAGTGTGAGTCAAAAGAGCAATGATCGAAAGCAG 600
 DB 601 AATAAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGCTTGTGATCGAAGAGATAAT 660
 QY 601 AATAAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGCTTGTGATCGAAGAGATAAT 660
 DB 601 AATAAGTGTGTTCAAAAAGTTCTTTGGCAGCTTGGATGCTTGTGATCGAAGAGATAAT 660
 QY 661 GAACCAAGAGGAGTGTGCTTGTGCAAGCTTGTGCAAGCAGCTGTGCTCATGATG 720
 DB 661 GAACCAAGAGGAGTGTGCTTGTGCAAGCTTGTGCAAGCAGCTGTGCTCATGATG 720
 QY 661 GAACCAAGAGGAGTGTGCTTGTGCAAGCTTGTGCAAGCAGCTGTGCTCATGATG 720
 DB 721 TATTAACCTGCAACCTTCTAAGTCAATGAGAGCAAAAGAGAGTAGGCAATTAAGCTC 780
 QY 721 TATTAACCTGCAACCTTCTAAGTCAATGAGAGCAAAAGAGAGTAGGCAATTAAGCTC 780
 DB 721 TATTAACCTGCAACCTTCTAAGTCAATGAGAGCAAAAGAGAGTAGGCAATTAAGCTC 780
 QY 781 CATAGAGAAATGGCTGAAGTTTAAAGAGATCTACAGATGAGAGGCGCATCTGTCATT 840
 DB 781 CATAGAGAAATGGCTGAAGTTTAAAGAGATCTACAGATGAGAGGCGCATCTGTCATT 840
 QY 781 CATAGAGAAATGGCTGAAGTTTAAAGAGATCTACAGATGAGAGGCGCATCTGTCATT 840
 DB 841 CTTTGTGTGAGATTTATGGGCAAGTGAAGGTTTGGGTGAGAGATGAAGAAAGTTATGAA 900
 QY 841 CTTTGTGTGAGATTTATGGGCAAGTGAAGGTTTGGGTGAGAGATGAAGAAAGTTATGAA 900
 DB 841 CTTTGTGTGAGATTTATGGGCAAGTGAAGGTTTGGGTGAGAGATGAAGAAAGTTATGAA 900

QY 901 ATTGCTGGCTTCGATATGATGATATCATGATACTAGACTGATTAAGAGCATTAACGAT 960
 DB 901 ATTGCTGGCTTCGATATGATGATATCATGATACTAGACTGATTAAGAGCATTAACGAT 960
 QY 961 AAGTGGCTGGAGATGGGACCAACAGAGCTGTGGAATCTTGCACAGATTAAGAGCTCTT 1020
 DB 961 AAGTGGCTGGAGATGGGACCAACAGAGCTGTGGAATCTTGCACAGATTAAGAGCTCTT 1020
 QY 1021 AGAGCAAGCCATTCCTATATGCTCAGAGGCGCATCGTGGAGATGATGTTTATG 1080
 DB 1021 AGAGCAAGCCATTCCTATATGCTCAGAGGCGCATCGTGGAGATGATGTTTATG 1080
 QY 1021 AGAGCAAGCCATTCCTATATGCTCAGAGGCGCATCGTGGAGATGATGTTTATG 1080
 DB 1081 AGAGATGCACTGGCTATTTTGAAGCGCAACGCTTCACCGGAGATGATGATGAGG 1140
 QY 1081 AGAGATGCACTGGCTATTTTGAAGCGCAACGCTTCACCGGAGATGATGATGAGG 1140
 DB 1081 AGAGATGCACTGGCTATTTTGAAGCGCAACGCTTCACCGGAGATGATGATGAGG 1140
 QY 1141 TTGATTAAGATTTGCTGGGGTCAAGAGCGCATGATTTTCTGAGAGTGTGCCAATG 1200
 DB 1141 TTGATTAAGATTTGCTGGGGTCAAGAGCGCATGATTTTCTGAGAGTGTGCCAATG 1200
 QY 1141 TTGATTAAGATTTGCTGGGGTCAAGAGCGCATGATTTTCTGAGAGTGTGCCAATG 1200
 DB 1201 TATGCTTCTGTCAGACAGAGCAAGATCTGAGATATTCATCAACACTCTCAAGCAAA 1260
 QY 1201 TATGCTTCTGTCAGACAGAGCAAGATCTGAGATATTCATCAACACTCTCAAGCAAA 1260
 DB 1201 TATGCTTCTGTCAGACAGAGCAAGATCTGAGATATTCATCAACACTCTCAAGCAAA 1260
 QY 1261 ACAAGCTGAATTCGATGAAATCAATACAGAGATGTTTAAAGAGCTGAGAGCAG 1320
 DB 1261 ACAAGCTGAATTCGATGAAATCAATACAGAGATGTTTAAAGAGCTGAGAGCAG 1320
 QY 1261 ACAAGCTGAATTCGATGAAATCAATACAGAGATGTTTAAAGAGCTGAGAGCAG 1320
 DB 1321 ATCTCTGAGCAATTCAGAGCTGAACTATCTTAAGAACAAAGCTCTCAAAACAGAAC 1380
 QY 1321 ATCTCTGAGCAATTCAGAGCTGAACTATCTTAAGAACAAAGCTCTCAAAACAGAAC 1380
 DB 1321 ATCTCTGAGCAATTCAGAGCTGAACTATCTTAAGAACAAAGCTCTCAAAACAGAAC 1380
 QY 1381 CACGCAAGGCTTGTGAGGATCTCTGAAATTAAGAGCAAGAGCTGCTAAGACTGCA 1440
 DB 1381 CACGCAAGGCTTGTGAGGATCTCTGAAATTAAGAGCAAGAGCTGCTAAGACTGCA 1440
 QY 1381 CACGCAAGGCTTGTGAGGATCTCTGAAATTAAGAGCAAGAGCTGCTAAGACTGCA 1440
 DB 1441 GAGGATTAATCGATCGTGAACAGAGAACTTAAGATGAGAGTGAACAGAGGAGAG 1500
 QY 1441 GAGGATTAATCGATCGTGAACAGAGAACTTAAGATGAGAGTGAACAGAGGAGAG 1500
 DB 1441 GAGGATTAATCGATCGTGAACAGAGAACTTAAGATGAGAGTGAACAGAGGAGAG 1500
 QY 1501 AATGATGCAACAGAGATTTTCAATGATTCATCAACAGATTCATGAAGAGAGAC 1560
 DB 1501 AATGATGCAACAGAGATTTTCAATGATTCATCAACAGATTCATGAAGAGAGAC 1560
 QY 1501 AATGATGCAACAGAGATTTTCAATGATTCATCAACAGATTCATGAAGAGAGAC 1560
 DB 1561 GCAAGAGAGAGATTTGAGATGTTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1561 GCAAGAGAGAGATTTGAGATGTTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 DB 1561 GCAAGAGAGAGATTTGAGATGTTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 1621 CAGCAGCAACATTAATCCCTCTAGCAATGAGATTCGAAAGAGAGAGAGAGAG 1680
 DB 1621 CAGCAGCAACATTAATCCCTCTAGCAATGAGATTCGAAAGAGAGAGAGAGAGAG 1680
 QY 1621 CAGCAGCAACATTAATCCCTCTAGCAATGAGATTCGAAAGAGAGAGAGAGAGAG 1680
 DB 1681 TCAGCTTCAATGAGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 QY 1681 TCAGCTTCAATGAGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 DB 1681 TCAGCTTCAATGAGATTTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
 QY 1741 CTGATTAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 DB 1741 CTGATTAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 QY 1741 CTGATTAAGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
 DB 1801 TTTGATCTGAGAGAGAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 QY 1801 TTTGATCTGAGAGAGAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 DB 1801 TTTGATCTGAGAGAGAGATTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 QY 1861 CACAAATGAGATGATGA 1878
 DB 1861 CACAAATGAGATGATGA 1878

RESULT 2
 AAF25373
 ID AAF25373 standard; DNA; 3275 BP.

XX AC AAF25373;
XX DT 15-MAY-2001 (first entry)
XX DE Genomic sequence of the Arabidopsis SGS3 gene.
XX KM SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX KM viral resistance; resistance; fatty acid content; protein content; ss.
XX OS Arabidopsis thaliana.
XX FH Key Location/Qualifiers
XX FT exon 696..1658
FT intron /tag= a
FT intron 1659..1731
FT exon /tag= b
FT intron 1732..2023
FT intron /tag= c
FT intron 2024..2134
FT exon /tag= d
FT intron 2135..2379
FT intron /tag= e
FT intron 2380..2481
FT exon /tag= f
FT intron 2482..2648
FT intron /tag= g
FT intron 2649..2738
FT exon /tag= h
FT exon 2739..2949
FT exon /tag= i
XX PN WO200105951-AZ.
XX PD 25-JAN-2001.
XX PF 13-JUL-2000; 2000WO-FR002052.
XX PR 16-JUL-1999; 99FR-00009417.
XX PR 26-JAN-2000; 2000FR-00001006.
XX PA (AVET) AVENTIS CROPS SCIENCE SA.
XX PA (INRG) INST NAT RECH AGRONOMIQUE.
XX PI Beclin C, Elmayan T, Vaucheret H;
XX WP1; 2001-159529/16.
XX PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX PT resistance in plants and, when inhibited, for increasing transgene
XX PT expression.
XX PS Claim 1; Page 31-32; 36pp; French.
XX CC The present sequence represents the genomic sequence of the Arabidopsis
CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
CC inactivation (degradation of RNA) and for resistance to viruses.
CC Overexpression of SGS3 results in plants with increased resistance to
CC viruses, while inactivation of SGS3 in transgenic plants (e.g. by
CC expressing antisense RNA, by mutation or by homologous recombination)
CC increases the level of the transgene product. This product may e.g.
CC impart resistance (to herbicide, insects or pathogens), alter contents of
CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an
CC immunoglobulin or interferon
XX SQ Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;
Query Match 51.3%; Score 963; DB 4; Length 3275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGAGTTCTAGGCGCTGCTCAATGCTTAAGAAAAAGAAAGTTCAAGGTTGTTATAGGCTT 60
|||||

Db 696 ATGAGTTCTAGGCGCTGCTCAATGCTTAAGAAAAAGAAAGTTCAAGGTTGTTATAGGCTT 755
Qy 61 GAGTTGAACAGTTGGTTCAAGTTTGGCAGGAGCAGAGCTGGCTTCTTCAACAATGAT 120
Db 756 GAGTTGAACAGTTGGTTCAAGTTTGGCAGGAGCAGAGCTGGCTTCTTCAACAATGAT 815
Qy 121 GGAGGAGAGTGGAGGAGTCAATTTCAAGAGAAACAAGAAACAACAGAAACAATTCTGGA 180
Db 816 GGAGGAGAGTGGAGGAGTCAATTTCAAGAGAAACAAGAAACAACAGAAACAATTCTGGA 875
Qy 181 AAAACTTGGGTTTCTCAAGATTCGAATCTCTTAAGCTTGGGTTGTCAGACGAAGG 240
Db 876 AAAACTTGGGTTTCTCAAGATTCGAATCTCTTAAGCTTGGGTTGTCAGACGAAGG 935
Qy 241 AGAGGTAGCAACGTATCTGGAGAGAGAAACAATGATCCGGAGAGGTAAACGCAATGGT 300
Db 936 AGAGGTAGCAACGTATCTGGAGAGAGAAACAATGATCCGGAGAGGTAAACGCAATGGT 995
Qy 301 CGGGGCATTCAAGCTTAACATATCTGGTGGGAGCAGGCGTTGAGCAGAAATATGATTAAC 360
Db 996 CGGGGCATTCAAGCTTAACATATCTGGTGGGAGCAGGCGTTGAGCAGAAATATGATTAAC 1055
Qy 361 AACTTTGGGCAACCCCACTGTATCTGCCCTCTTTGAAAGAGATGAATTTGGCAG 420
Db 1056 AACTTTGGGCAACCCCACTGTATCTGCCCTCTTTGAAAGAGATGAATTTGGCAG 1115
Qy 421 GCAGAGAGAGGTTCTGCTCAGCAGCAGCTGTGAGAGGTTTCTGAGTGGAGATGAT 480
Db 1116 GCAGAGAGAGGTTCTGCTCAGCAGCAGCTGTGAGAGGTTTCTGAGTGGAGATGAT 1175
Qy 481 GTGATTAATGCTTCTGAGAGAGAGATATTCGATGCTTTGATGATTTCTGATGACGAC 540
Db 1176 GTGATTAATGCTTCTGAGAGAGAGATATTCGATGCTTTGATGATTTCTGATGACGAC 1235
Qy 541 CTGGCAAGTATGATTAATGATCTCGATGTGAGTCAAAAGAGCATGATCACGAAACGAG 600
Db 1236 CTGGCAAGTATGATTAATGATCTCGATGTGAGTCAAAAGAGCATGATCACGAAACGAG 1295
Qy 601 AATAAGTGTTCAAAGATTTCTTGGCAGCTTGGATGATCTGTGATGAGCAGATTAAT 660
Db 1296 AATAAGTGTTCAAAGATTTCTTGGCAGCTTGGATGATCTGTGATGAGCAGATTAAT 1355
Qy 661 GAACACAGAGGAGGAGTGCATTTGTCAGCTTGTCAAGACGACCTGTGCCATCGATTGG 720
Db 1356 GAACACAGAGGAGGAGTGCATTTGTCAGCTTGTCAAGACGACCTGTGCCATCGATTGG 1415
Qy 721 TATTAACCTGACACCTCTCTAAGCTTCATGCTGATGCGAGGACAAAGAGCTAAGCTC 780
Db 1416 TATTAACCTGACACCTCTCTAAGCTTCATGCTGATGCGAGGACAAAGAGCTAAGCTC 1475
Qy 781 CATAGAGAAATGGCTGGAAGTTTAAAGAAAGATCTACAGATGAGGCGCATCTGTCAAT 840
Db 1476 CATAGAGAAATGGCTGGAAGTTTAAAGAAAGATCTACAGATGAGGCGCATCTGTCAAT 1535
Qy 841 CTTGTGTGAGATTTATGAGCAGTGAAGGTTTGGTGGAGATGAAGAAAGATTAATGA 900
Db 1536 CTTGTGTGAGATTTATGAGCAGTGAAGGTTTGGTGGAGATGAAGAAAGATTAATGA 1595
Qy 901 ATTGTCTGAGCTTCAGATGATCATCATGATTAATGATGATGATTAAGCAGATTAACAT 960
Db 1596 ATTGTCTGAGCTTCAGATGATCATCATGATTAATGATGATGATTAAGCAGATTAACAT 1655
Qy 961 AAG 963
Db 1656 AAG 1658
RESULT 3
AAF25372/C
ID AAF25372 standard; DNA; 27 BP.
XX AAF25372;
XX

```

DT      15-MAY-2001 (first entry)
XX
DE      PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
XX      SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX      viral resistance; resistance; fatty acid content; protein content;
XX      PCR primer; ss.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200105951-A2.
XX
XX      25-JAN-2001.
XX
XX      13-JUL-2000; 2000WO-FR002052.
XX
XX      16-JUL-1999; 99FR-00009417.
XX      26-JAN-2000; 2000FR-00001006.
XX
XX      (AVET ) AVENTIS CROPSCHIENCE SA.
XX      (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX      Beclin C, Elmayer T, Vaucheret H;
XX
XX      WPI; 2001-159529/16.
XX
XX      New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX      resistance in plants and, when inhibited, for increasing transgene
XX      expression.
XX
XX      Example 1; Page 22; 36pp; French.
XX
XX      PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
XX      thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
XX      transcriptional inactivation (degradation of RNA) and for resistance to
XX      viruses. Overexpression of SGS3 results in plants with increased
XX      resistance to viruses, while inactivation of SGS3 in transgenic plants
XX      (e.g. by expressing antisense RNA, by mutation or by homologous
XX      recombination) increases the level of the transgene product. This product
XX      may e.g. impart resistance (to herbicide, insects or pathogens), alter
XX      contents of essential fatty acids or proteins, or is pharmaceutically
XX      active, e.g. an immunoglobulin or interferon
XX
XX      Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;
XX
XX      Query Match      1.3%; Score 24; DB 4; Length 27;
XX      Best local Similarity 100.0%; Pred. No. 0.71;
XX      Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX      QY      1855 GGCCTTCACATGAGATGATTGA 1878
XX      DB      27 GGCCTTCACATGAGATGATTGA 4
XX
XX      RESULT 4
XX      AAF25371
XX      ID      AAF25371 standard; DNA; 23 BP.
XX
XX      AC      AAF25371;
XX
XX      DT      15-MAY-2001 (first entry)
XX
XX      DE      PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
XX      SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX      viral resistance; resistance; fatty acid content; protein content;
XX      PCR primer; ss.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200105951-A2.
XX
XX      25-JAN-2001.
XX
XX      13-JUL-2000; 2000WO-FR002052.
XX
XX      16-JUL-1999; 99FR-00009417.
XX      26-JAN-2000; 2000FR-00001006.
XX
XX      (AVET ) AVENTIS CROPSCHIENCE SA.
XX      (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX      Beclin C, Elmayer T, Vaucheret H;
XX
XX      WPI; 2001-159529/16.
XX
XX      New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX      resistance in plants and, when inhibited, for increasing transgene
XX      expression.
XX
XX      Example 1; Page 22; 36pp; French.
XX
XX      PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
XX      thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
XX      transcriptional inactivation (degradation of RNA) and for resistance to
XX      viruses. Overexpression of SGS3 results in plants with increased
XX      resistance to viruses, while inactivation of SGS3 in transgenic plants
XX      (e.g. by expressing antisense RNA, by mutation or by homologous
XX      recombination) increases the level of the transgene product. This product
XX      may e.g. impart resistance (to herbicide, insects or pathogens), alter
XX      contents of essential fatty acids or proteins, or is pharmaceutically
XX      active, e.g. an immunoglobulin or interferon
XX
XX      Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;
XX
XX      Query Match      1.3%; Score 24; DB 4; Length 27;
XX      Best local Similarity 100.0%; Pred. No. 0.71;
XX      Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX      QY      1855 GGCCTTCACATGAGATGATTGA 1878
XX      DB      27 GGCCTTCACATGAGATGATTGA 4
XX
XX      RESULT 4
XX      AAF25371
XX      ID      AAF25371 standard; DNA; 23 BP.
XX
XX      AC      AAF25371;
XX
XX      DT      15-MAY-2001 (first entry)
XX
XX      DE      PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
XX      SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX      viral resistance; resistance; fatty acid content; protein content;
XX      PCR primer; ss.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200105951-A2.
XX
XX      25-JAN-2001.
XX
XX      13-JUL-2000; 2000WO-FR002052.
XX
XX      16-JUL-1999; 99FR-00009417.
XX      26-JAN-2000; 2000FR-00001006.
XX
XX      (AVET ) AVENTIS CROPSCHIENCE SA.
XX      (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX      Beclin C, Elmayer T, Vaucheret H;
XX
XX      WPI; 2001-159529/16.
XX
XX      New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX      resistance in plants and, when inhibited, for increasing transgene
XX      expression.
XX
XX      Example 1; Page 22; 36pp; French.
XX
XX      PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
XX      thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
XX      transcriptional inactivation (degradation of RNA) and for resistance to
XX      viruses. Overexpression of SGS3 results in plants with increased
XX      resistance to viruses, while inactivation of SGS3 in transgenic plants
XX      (e.g. by expressing antisense RNA, by mutation or by homologous
XX      recombination) increases the level of the transgene product. This product
XX      may e.g. impart resistance (to herbicide, insects or pathogens), alter
XX      contents of essential fatty acids or proteins, or is pharmaceutically
XX      active, e.g. an immunoglobulin or interferon
XX
XX      Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;
XX
XX      Query Match      1.3%; Score 24; DB 4; Length 27;
XX      Best local Similarity 100.0%; Pred. No. 0.71;
XX      Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX      QY      1855 GGCCTTCACATGAGATGATTGA 1878
XX      DB      27 GGCCTTCACATGAGATGATTGA 4
XX
XX      RESULT 4
XX      AAF25371
XX      ID      AAF25371 standard; DNA; 23 BP.
XX
XX      AC      AAF25371;
XX
XX      DT      15-MAY-2001 (first entry)
XX
XX      DE      PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
XX      SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX      viral resistance; resistance; fatty acid content; protein content;
XX      PCR primer; ss.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200105951-A2.
XX
XX      25-JAN-2001.
XX
XX      13-JUL-2000; 2000WO-FR002052.
XX
XX      16-JUL-1999; 99FR-00009417.
XX      26-JAN-2000; 2000FR-00001006.
XX
XX      (AVET ) AVENTIS CROPSCHIENCE SA.
XX      (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX      Beclin C, Elmayer T, Vaucheret H;
XX
XX      WPI; 2001-159529/16.
XX
XX      New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX      resistance in plants and, when inhibited, for increasing transgene
XX      expression.
XX
XX      Example 1; Page 22; 36pp; French.
XX
XX      PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
XX      thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
XX      transcriptional inactivation (degradation of RNA) and for resistance to
XX      viruses. Overexpression of SGS3 results in plants with increased
XX      resistance to viruses, while inactivation of SGS3 in transgenic plants
XX      (e.g. by expressing antisense RNA, by mutation or by homologous
XX      recombination) increases the level of the transgene product. This product
XX      may e.g. impart resistance (to herbicide, insects or pathogens), alter
XX      contents of essential fatty acids or proteins, or is pharmaceutically
XX      active, e.g. an immunoglobulin or interferon
XX
XX      Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;
XX
XX      Query Match      1.3%; Score 24; DB 4; Length 27;
XX      Best local Similarity 100.0%; Pred. No. 0.71;
XX      Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX      QY      1855 GGCCTTCACATGAGATGATTGA 1878
XX      DB      27 GGCCTTCACATGAGATGATTGA 4
XX
XX      RESULT 4
XX      AAF25371
XX      ID      AAF25371 standard; DNA; 23 BP.
XX
XX      AC      AAF25371;
XX
XX      DT      15-MAY-2001 (first entry)
XX
XX      DE      PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
XX      SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX      viral resistance; resistance; fatty acid content; protein content;
XX      PCR primer; ss.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200105951-A2.
XX
XX      25-JAN-2001.
XX
XX      13-JUL-2000; 2000WO-FR002052.
XX
XX      16-JUL-1999; 99FR-00009417.
XX      26-JAN-2000; 2000FR-00001006.
XX
XX      (AVET ) AVENTIS CROPSCHIENCE SA.
XX      (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX      Beclin C, Elmayer T, Vaucheret H;
XX
XX      WPI; 2001-159529/16.
XX
XX      New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX      resistance in plants and, when inhibited, for increasing transgene
XX      expression.
XX
XX      Example 1; Page 22; 36pp; French.
XX
XX      PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
XX      thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
XX      transcriptional inactivation (degradation of RNA) and for resistance to
XX      viruses. Overexpression of SGS3 results in plants with increased
XX      resistance to viruses, while inactivation of SGS3 in transgenic plants
XX      (e.g. by expressing antisense RNA, by mutation or by homologous
XX      recombination) increases the level of the transgene product. This product
XX      may e.g. impart resistance (to herbicide, insects or pathogens), alter
XX      contents of essential fatty acids or proteins, or is pharmaceutically
XX      active, e.g. an immunoglobulin or interferon
XX
XX      Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;
XX
XX      Query Match      1.3%; Score 24; DB 4; Length 27;
XX      Best local Similarity 100.0%; Pred. No. 0.71;
XX      Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX      QY      1855 GGCCTTCACATGAGATGATTGA 1878
XX      DB      27 GGCCTTCACATGAGATGATTGA 4
XX
XX      RESULT 4
XX      AAF25371
XX      ID      AAF25371 standard; DNA; 23 BP.
XX
XX      AC      AAF25371;
XX
XX      DT      15-MAY-2001 (first entry)
XX
XX      DE      PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
XX      SGS3 gene; post-transcriptional inactivation; RNA degradation;
XX      viral resistance; resistance; fatty acid content; protein content;
XX      PCR primer; ss.
XX
XX      Arabidopsis thaliana.
XX
XX      WO200105951-A2.
XX
XX      25-JAN-2001.
XX
XX      13-JUL-2000; 2000WO-FR002052.
XX
XX      16-JUL-1999; 99FR-00009417.
XX      26-JAN-2000; 2000FR-00001006.
XX
XX      (AVET ) AVENTIS CROPSCHIENCE SA.
XX      (INRG ) INST NAT RECH AGRONOMIQUE.
XX
XX      Beclin C, Elmayer T
```

Pf	XX
Pf	XX
PR	13-JUL-2000; 2000MO-FR002052.
FR	16-JUL-1999; 99FR-00009417.
PA	26-JAN-2000; 2000FR-00001006.
PI	(AVET) AVENTIS CROPS&CIE SA.
PP	(INRG) INST NAT RECH AGROMIQUE.
PT	Beclin C, Elmayan T, Vaucheret H;
PU	WPI; 2001-159529/16.
PV	New SGS3 gene from Arabidopsis thaliana, useful for increasing virus resistance in plants and, when inhibited, for increasing transgene expression.
QD	Example 1; Page 22; 36pp; French.
QC	PCR primers AAPF23J71-72 were used to amplify cDNA encoding an Arabidopsis chlamydomonas polyprotein. The SGS3 gene is essential for post-transcriptional inactivation of SGS3 results in plants with increased viruses. Overexpression of SGS3 results in plants with increased resistance to viruses, while inactivation of SGS3 in transgenic plants (e.g.) by expressing antisense RNA, by mutation or by homologous recombination) increases the level of the transgene product. This product may e.g., impart resistance (to herbicide, insects or pathogens), alter contents of essential fatty acids or proteins, or is pharmaceutically active, e.g., an immunoglobulin or interferon
QS	Sequence 23 BP; 6 A; 4 G; 7 T; 0 U; 0 Other;
QU	Query Match 1.1%; Score 20; DB 4; Length 23; Best Local Similarity 100.0%; Pred. No. 60; Indels 0; Gaps 0 Matches 20; Conservative 0; Mismatches 0
RH	1 ATGAGTTCTAAGCGTGTCC 20 4 ATGAGTTCTAAGCGTGTCC 23
RE	RESULT 5
RF	AAPF12773
RI	ID AAPF12773 standard; CDNA; 596 BP.
RK	AAAF12773:
RL	DAT 13-MAR-2001 (first entry)
RM	Aspergillus oryzae EST SEQ ID NO:5296.
RN	Multiplex gene expression; filamentous fungal cell; EST;
RP	expressed sequence tag; Fusarium venenatum; Aspergillus niger;
RR	Aspergillus oryzae; Trichoderma reesei; identification; recombinant;
RS	culture condition; environmental stress; spore morphogenesis;
RT	metabolic pathway engineering; metabolic pathway engineering; ss.
RU	Aspergillus oryzae.
RV	WO2000056762-A2.
RX	28-SEP-2000.
RY	22-MAR-2000; 2000MO-US007781.
SZ	22-MAR-1999; 99US-00273623.
TB	(NOVO) NOVO NORDISK BIOTECH INC.
TD	(NOVO) NOVO NORDISK AS.
TE	Berka RM, Rey MM, Shuster JR, Kaupinen S, Clausen IG, Olsen PB;
TF	WPI; 2000-594572/56.

CC Neisseria meningitidis DNA sequences; and AAA81322 to AAA81452 represent
 CC Neisseria meningitidis MenB polymucloide ORF sequences, which are all
 CC used in the amplification of the present invention. The nucleic acid
 CC sequences, protein sequences, and antibodies against them, can be used in
 CC the manufacture of a composition. The composition can be used as a
 CC medicament (or in the manufacture of a medicament) for treating,
 CC preventing or diagnosing infection due to Neisserial bacteria. For
 CC example, some of the identified proteins could be components of vaccines
 CC against Meningococcus B; against all serotypes; and/or against all
 CC pathogenic Neisseriae. Identification of sequences from the bacterium
 CC will also facilitate production of biological probes, particularly
 CC organism-specific probes. Attempts to make efficacious Meningococcus B
 CC vaccines have also been tried but none have successfully overcome
 CC antigenic variability. The provision of further, complete sequences may
 CC provide an opportunity to identify secreted or surface exposed proteins
 CC that may be presumed targets for the immune system and which are not
 CC antigenically variable or at least more conserved than other more
 CC variable regions

XX
 XX Sequence 47475 BP; 11600 A; 13758 C; 12290 G; 9826 T; 0 U; 1 Other;
 SO

Query Match 1.1%; Score 20; DB 3; Length 47475;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 GAACAGTTGGTTCAGGTTT 86
 DB 23354 GAACAGTTGGTTCAGGTTT 23335

RESULT 8
 AAA81489 5/c
 Continuation (6 of 9) of AAA81489 from base 500001 (N. meningitidis partial DNA sequence
 WP Sequence split into 9 fragments LOCUS AAA81489 Accession Aaa81489
 WP Fragment Name Begin End
 WP AAA81489_0 1 110000
 WP AAA81489_1 100001 210000
 WP AAA81489_2 200001 310000
 WP AAA81489_3 300001 410000
 WP AAA81489_4 400001 510000
 WP AAA81489_5 500001 610000
 WP AAA81489_6 600001 710000
 WP AAA81489_7 700001 810000
 WP AAA81489_8 800001 837096

Query Match 1.1%; Score 20; DB 3; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 67 GAACAGTTGGTTCAGGTTT 86
 DB 32595 GAACAGTTGGTTCAGGTTT 32576

RESULT 9
 ABO88198/c
 ID ABO88198 standard; cDNA; 154902 BP.
 AC ABO88198;
 XX
 XX 18-SEP-2002 (first entry)
 DT
 XX Human osteoblast differentiation related cDNA SEQ ID NO 105.
 DE
 XX Human; osteoblast; stem cell differentiation; bone tissue deposition;
 XX osteoporosis; osteopathic; ss.
 KM
 XX Homo sapiens.
 OS
 XX WO200250301-A2.
 PN
 XX 27-JUN-2002.
 PD

XX
 XX 18-DEC-2001; 2001MO-US048276.
 PF
 XX 18-DEC-2000; 2000US-0255882P.
 PR
 XX 24-APR-2001; 2001US-0285691P.
 PR
 XX (GENE-) GENE LOGIC INC.
 PA (PROC) PROCTER & GAMBLE CO.
 XX
 XX JI D, Axelrod DW, Cook JS, Jaiswal N, Einstein R, Houghton A;
 PI Mertz L;
 XX WPI; 2002-557663/59.
 DR
 XX Use of genes and their expression profiles associated with osteoblast
 PT differentiation for screening modulators bone formation, for diagnosing
 PT or treating e.g. osteoporosis, or as markers for the differentiation
 PT process.
 XX
 XX Claim 1; SEQ ID NO 105; 78bp + Sequence Listing; English.
 PS
 XX The invention relates to genes and their expression profiles are used
 CC for: (a) screening modulators of precursor stem cell differentiation into
 CC osteoblasts, or bone tissue deposition; (b) diagnosing abnormal
 CC deposition of bone tissue, abnormal rate of osteoblast formation or
 CC osteoporosis; or (c) treating or monitoring treatment of the conditions
 CC cited in (b) or monitoring the progression of bone tissue deposition.
 CC Specific conditions include postmenopausal osteoporosis, glucocorticoid
 CC induced abnormalities in bone formation or bone loss, conditions that
 CC involve altered bone metabolism (e.g. idiopathic juvenile osteoporosis),
 CC skeletal disease linked to breast cancer, mastocytosis, Fanconi syndrome
 CC or fibrous dysplasia. The present sequence is that of an osteoblast
 CC differentiation associated cDNA marker of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

XX
 XX Sequence 154902 BP; 43917 A; 31458 C; 32848 G; 46679 T; 0 U; 0 Other;
 SO

Query Match 1.1%; Score 20; DB 6; Length 154902;
 Best Local Similarity 100.0%; Pred. No. 54;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1712 TGGAGAGTTTGTGGAAGG 1731
 DB 128153 TGGAGAGTTTGTGGAAGG 128134

RESULT 10
 AAF21612/c
 ID AAF21612 standard; DNA; 349980 BP.
 AC AAF21612;
 XX
 XX 13-MAR-2001 (first entry)
 DT
 XX Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
 DE
 XX Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;
 KM diagnosis; antigen; detection; infection; gene therapy; antibacterial;
 KM ds.
 XX
 XX Neisseria meningitidis.
 OS
 XX WO2000066791-A1.
 FN
 XX 09-NOV-2000.
 PD
 XX 08-MAR-2000; 2000MO-US005928.
 PF
 XX 30-APR-1999; 99US-0132068P.
 XX 08-OCT-1999; 99MO-US023573.
 PR

PR 28-FEB-2000; 2000GB-00004695.
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tetelín H, Venter JC,
PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V,
XX Rappuoli R, Frazer CM, Grandi G;
XX WPI: 2000-647603/62.
XX
PT *Neisseria meningitidis* B full length genome sequence and open reading
frames are used to detect, treat and prevent *Neisseria* infections.
XX
PS Claim 7; Appendix A; 692pp; English.
XX
XX The present invention describes the full length genome of *Neisseria*
meningitidis B (NMB). The sequences in AAF21544 and AAF21607 to AAF21613
represent fragments of the NMB genomic sequence, as the sequence was too
long to go in a record on its own it was split into 8 sequences which
overlap each other at the beginning and end of each sequence by 49980 bp
(i.e. the last 49980 bp of AAF21544 is repeated at the beginning of
AAF21607, the last 49980 bp of AAF21607 are repeated at the beginning of
AAF21608, and so on). AAF21545 to AAF21588 encode the *Neisseria* proteins
given in AAB58550 to AAB58593, and AAF21589 to AAF21606 represent PCR
primers which are used in the exemplification of the present invention.
CC The NMB genome and fragments from it have antibacterial activity, and can
be used in vaccines and gene therapy. *Neisseria* nucleic acids, proteins
and/or antibodies which binds to the proteins can be used in compositions
for treating or preventing infection due to *Neisseria* bacteria or as a
CC diagnostic reagent for detecting the presence of *Neisseria* bacteria or
of antibodies raised to *Neisseria* bacteria. Computers, computer memory,
CC computer storage medium or computer databases can be used in a search to
identify open reading frames (ORFs) or coding sequences within the NMB
CC genome. The DNA sequences provide further opportunities to find antigenic
CC or immunogenic proteins which are more effective in vaccines than the
outer membrane proteins currently used
XX
SQ Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 U; 0 Other;
XX
Query Match 1.1%; Score 20; DB 3; Length 349980;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 67 GAACAGTGGTCAAGGTTT 86
Db 167824 GAACAGTGGTCAAGGTTT 167805
XX
RESULT 11
AAA68016/C
ID AAA68016 standard; DNA; 481 BP.
XX
AC AAA68016;
XX
DT 24-OCT-2000 (first entry)
XX
DE Eucalyptus grandis CCR nucleotide sequence SEQ ID NO:109.
XX
KW Plant; lignin; lignin biosynthetic pathway; Eucalyptus grandis;
KW Pinus radiata; Monterey pine; ds.
XX
OS Eucalyptus grandis.
XX
FN WO200022099-A1.
XX
PD 20-APR-2000.
XX
PF 06-OCT-1999; 99WO-NZ000168.
XX
PR 09-OCT-1998; 98US-00169789.
PR 14-JUL-1999; 99US-0143811P.
XX

PA (GENE-) GENESIS RES & DEV CORP LTD.
XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Bloksberg LN, Havukkala IT;
XX
DR WPI: 2000-317962/27.
XX
PT Novel polynucleotide encoding enzymes involved in lignin-biosynthetic
pathway useful for producing transgenic plants especially eucalyptus and
PT pine species having altered lignin content, composition and structure.
XX
PS Claim 1; Page 91-92; 213pp; English.
XX
XX The present invention describes isolated polynucleotides and proteins
encoding and representing the enzymes cinnamate 4-hydroxylase (CAH),
CC coumarate 3-hydroxylase (CH), phenolase (PNU), O-methyl transferase
CC (OMT), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl-CoA reductase
CC (CCR), phenylalanine ammonia-lyase (PAL), 4-coumarate:CoA ligase (4CL),
CC coniferol glucosyl transferase (CGT), coniferin beta-glucosidase (CBG),
CC laccase, peroxidase, ferulate-5-hydroxylase (F5H), alpha-amylase, caffeic
CC acid methyl transferase, caffeoyl CoA methyl transferase, coumarate CoA
CC ligase, cytochrome P450 1X1A, diphenol oxidase, flavanol glucosyl
CC transferase, flavonoid hydroxylase, and isoflavone reductase, which are
CC involved in the lignin biosynthetic pathway. The polynucleotides can be
CC used for modulating lignin content, lignin composition and the structure
CC of a plant, especially eucalyptus and pine species, and for modifying the
CC activity of an enzyme involved in lignin biosynthetic pathway, and for
CC producing a plant having altered lignin content, composition and
CC structure. They can be used for designing probes and primers useful for
CC detecting similar DNA and RNA sequences in any organism and for PCR
CC amplification. The lignin content can be efficiently modified using the
CC polynucleotides. AAB67908 to AAB68201 and AAB16341 to AAB16449 represent
CC polynucleotide and protein sequences used in the exemplification of the
present invention
XX
SQ Sequence 481 BP; 122 A; 118 C; 111 G; 130 T; 0 U; 0 Other;
XX
Query Match 1.0%; Score 19; DB 3; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 TGGTCAAGGTTTGGCAGG 92
Db 470 TGGTCAAGGTTTGGCAGG 452
XX
RESULT 12
ADD41766/C
ID ADD41766 standard; DNA; 481 BP.
XX
AC ADD41766;
XX
DT 15-JAN-2004 (first entry)
XX
DE Cinnamoyl-CoA reductase DNA #21.
XX
KW ds; plant; lignin content; lignin composition; lignin structure;
KW lignin biosynthesis; gene.
XX
OS Eucalyptus grandis.
XX
PN US2003131373-A1.
XX
PD 10-JUL-2003.
XX
PF 18-JUN-2002; 2002US-00174693.
XX
PR 11-SEP-1996; 96US-00713000.
PR 21-NOV-1997; 97US-00975316.
PR 09-OCT-1998; 98US-00169789.
PR 14-JUL-1999; 99US-0143833P.
PR 12-JUL-2000; 2000US-00615192.
XX

PA (BLOCK) BLOKSBERG L N.
PA (HAYU/) HAYUKKALA I.
XX
XX
PI Blosberg LN, Havukkala I;
XX WPI; 2003-829606/77.
XX
XX New polynucleotide associated with the lignin biosynthetic pathway, for
PT modulating lignin content, composition and structure of plants, or
PT producing a plant with altered lignin content, composition and structure.
XX
PS Claim 1, SEQ ID NO 109; 206pp; English.
XX
XX The invention relates to an isolated polynucleotide. The polynucleotides,
CC polypeptides and genetic constructs are useful for modulating lignin
CC content, composition and structure of plants, or for producing a plant
CC having altered lignin content, composition and structure. The
CC polynucleotides are also useful in genome mapping, physical mapping,
CC positional cloning of genes, or as non-disruptive tags for marking
CC organisms, particularly plants. The present sequence represents a
CC polynucleotide associated with the lignin biosynthetic pathway.
XX
XX Sequence 481 BP; 122 A; 118 C; 111 G; 130 T; 0 U; 0 Other;
SQ
Query Match 1.0%; Score 19; DB 9; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 74 TGGTTCAAGGTTGGCAGG 92
Db 470 TGGTTCAAGGTTGGCAGG 452
RESULT 13
AAK64246
ID AAK64246 standard; CDNA; 484 BP.
XX
XX AAK64246;
AC
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:9306.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytosolic; gene therapy; vaccine; metastasis; ss.
XX
XX Homo sapiens.
OS
XX
XX WO200157182-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US001354.
PF
XX
XX 31-JAN-2000; 2000US-0179065P.
PR
XX 04-FEB-2000; 2000US-0180628P.
PR
XX 24-FEB-2000; 2000US-0184664P.
PR
XX 02-MAR-2000; 2000US-0186550P.
PR
XX 16-MAR-2000; 2000US-0189874P.
PR
XX 17-MAR-2000; 2000US-0190076P.
PR
XX 18-APR-2000; 2000US-0198123P.
PR
XX 19-MAY-2000; 2000US-0205515P.
PR
XX 07-JUN-2000; 2000US-0214886P.
PR
XX 28-JUN-2000; 2000US-0215135P.
PR
XX 30-JUN-2000; 2000US-0216647P.
PR
XX 07-JUL-2000; 2000US-0216880P.
PR
XX 11-JUL-2000; 2000US-0217487P.
PR
XX 11-JUL-2000; 2000US-0217496P.
PR
XX 14-JUL-2000; 2000US-0220963P.
PR
XX 26-JUL-2000; 2000US-0220964P.
PR
XX 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.
PR 14-AUG-2000; 2000US-0225213P.
PR 14-AUG-2000; 2000US-0225214P.
PR 14-AUG-2000; 2000US-0225266P.
PR 14-AUG-2000; 2000US-0225267P.
PR 14-AUG-2000; 2000US-0225268P.
PR 14-AUG-2000; 2000US-0225270P.
PR 14-AUG-2000; 2000US-0225447P.
PR 14-AUG-2000; 2000US-0225757P.
PR 14-AUG-2000; 2000US-0225758P.
PR 14-AUG-2000; 2000US-0225759P.
PR 14-AUG-2000; 2000US-0226279P.
PR 18-AUG-2000; 2000US-0226281P.
PR 22-AUG-2000; 2000US-0226868P.
PR 22-AUG-2000; 2000US-0227182P.
PR 22-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 01-SEP-2000; 2000US-0229509P.
PR 03-SEP-2000; 2000US-0229513P.
PR 05-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 06-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 08-SEP-2000; 2000US-0231968P.
PR 12-SEP-2000; 2000US-0231969P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 14-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 21-SEP-2000; 2000US-0234998P.
PR 25-SEP-2000; 2000US-0234999P.
PR 25-SEP-2000; 2000US-0235484P.
PR 26-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 27-SEP-2000; 2000US-0235837P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 29-SEP-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 02-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 13-OCT-2000; 2000US-0239938P.
PR 20-OCT-2000; 2000US-0241211P.
PR 20-OCT-2000; 2000US-0241212P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.
PR 08-NOV-2000; 2000US-0246477P.

PR 08-NOV-2000; 2000US-0246478P.
 PR 08-NOV-2000; 2000US-0246523P.
 PR 08-NOV-2000; 2000US-0246524P.
 PR 08-NOV-2000; 2000US-0246525P.
 PR 08-NOV-2000; 2000US-0246526P.
 PR 08-NOV-2000; 2000US-0246527P.
 PR 08-NOV-2000; 2000US-0246528P.
 PR 08-NOV-2000; 2000US-0246532P.
 PR 08-NOV-2000; 2000US-0246609P.
 PR 08-NOV-2000; 2000US-0246610P.
 PR 08-NOV-2000; 2000US-0246611P.
 PR 08-NOV-2000; 2000US-0246613P.
 PR 17-NOV-2000; 2000US-0249207P.
 PR 17-NOV-2000; 2000US-0249208P.
 PR 17-NOV-2000; 2000US-0249209P.
 PR 17-NOV-2000; 2000US-0249210P.
 PR 17-NOV-2000; 2000US-0249211P.
 PR 17-NOV-2000; 2000US-0249212P.
 PR 17-NOV-2000; 2000US-0249213P.
 PR 17-NOV-2000; 2000US-0249214P.
 PR 17-NOV-2000; 2000US-0249215P.
 PR 17-NOV-2000; 2000US-0249216P.
 PR 17-NOV-2000; 2000US-0249217P.
 PR 17-NOV-2000; 2000US-0249218P.
 PR 17-NOV-2000; 2000US-0249244P.
 PR 17-NOV-2000; 2000US-0249245P.
 PR 17-NOV-2000; 2000US-0249246P.
 PR 17-NOV-2000; 2000US-0249255P.
 PR 17-NOV-2000; 2000US-0249297P.
 PR 17-NOV-2000; 2000US-0249299P.
 PR 17-NOV-2000; 2000US-0249300P.
 PR 01-DEC-2000; 2000US-0250160P.
 PR 01-DEC-2000; 2000US-0250391P.
 PR 05-DEC-2000; 2000US-0251030P.
 PR 05-DEC-2000; 2000US-0251888P.
 PR 05-DEC-2000; 2000US-0256719P.
 PR 06-DEC-2000; 2000US-0251479P.
 PR 08-DEC-2000; 2000US-0251856P.
 PR 08-DEC-2000; 2000US-0251868P.
 PR 08-DEC-2000; 2000US-0251869P.
 PR 08-DEC-2000; 2000US-0251899P.
 PR 08-DEC-2000; 2000US-0251907P.
 PR 11-DEC-2000; 2000US-0254097P.
 PR 05-JAN-2001; 2001US-0259678P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-483426/52.
 DR P-PSDB; AAM91465.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and metastasis.
 PS
 PS
 XX Claim 1, SEQ ID NO 9306; 3071pp + Sequence Listing; English.
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting the
 CC nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK87694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169

CC represent sequences used in the exemplification of the present invention
 XX
 XX Sequence 484 BP; 179 A; 66 C; 106 G; 131 T; 0 U; 2 Other;
 SQ
 Query Match 1.0%; Score 19; DB 4; Length 484;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1767 GGAGACATGAGAGAGG 1785
 |||||
 DB 323 GGAGACATGAGAGAGG 341
 RESULT 14
 AAV23847/c
 ID AAV23847 standard; DNA; 495 BP.
 XX
 AC AAV23847;
 XX
 DT 31-JUL-1998 (first entry)
 XX
 DE Plant CCR enzyme DNA sequence.
 XX
 KW Lignin biosynthetic pathway; eucalyptus; pine; transgenic plant;
 KW lignin content; tree processing; cellulose fibre; ss.
 XX
 OS Eucalyptus grandis.
 XX
 PN WO9811205-A2.
 XX
 PD 19-MAR-1998.
 XX
 PF 10-SEP-1997; 97WO-NZ000112.
 XX
 PR 11-SEP-1996; 96US-00713000.
 XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLEET-) FLEETCHER CHALLENGE FORESTS LTD.
 XX
 PI Bloksberg LN, Grierson AW, Havukkala IU;
 DR WPI: 1998-207374/18.
 XX
 PT Sequences useful for modification of plant lignin content or structure -
 PT from Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) are
 PT associated with lignin biosynthesis pathway, useful e.g. in paper
 PT industry.
 PS
 PS Claim 1; Page 39; 82pp; English.
 PS
 XX This sequence represents a fragment of the CCR enzyme coding sequence. It
 CC is an example of a DNA sequence of the invention, which are from
 CC Eucalyptus grandis (eucalyptus) and Pinus radiata (pine) associated with
 CC the lignin biosynthesis pathway. Constructs containing the DNA sequences
 CC can be used to produce transgenic plants or plant cells, especially woody
 CC plants e.g. eucalyptus or pine species but also e.g. monocotyledons or
 CC dicotyledons; by stably incorporating the constructs into the plant
 CC genome. The lignin content or structure, or activity of a specific enzyme
 CC in the plant, can therefore be modulated. Reductions in lignin content or
 CC changes in composition are useful in tree processing for paper. High
 CC lignin content results in energy- and chemical-intensive separation
 CC methods in order to obtain the pure cellulose fibre required. Reductions
 CC in lignin content may also be useful for forage crops, whilst increases
 CC or changes in composition may be desirable to increase the mechanical
 CC strength of wood, change its colour or increase its resistance to rot.
 CC The sequences are also useful as probes to isolate DNA sequences encoding
 CC enzymes involved in the lignin biosynthesis pathway from other plant
 CC species
 XX
 SQ Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;
 Query Match 1.0%; Score 19; DB 2; Length 495;
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Search completed: March 30, 2004, 09:26:50
Job time : 655.082 secs

Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 74 TGGTCAAGTTGGCAGG 92
 |||||
 Db 484 TGGTCAAGTTGGCAGG 466

RESULT 15
 AA206848/c
 ID AA206848 standard; cDNA; 495 BP.

XX AA206848;
 AC

XX 09-NOV-1999 (first entry)
 DT

XX Eucalyptus cinamoyl-CoA reductase (CCR) partial cDNA 2.
 DE

XX Lignin; biosynthesis; forage crop; wood; paper production;
 KM

XX transgenic plant; ss.
 KW

XX Eucalyptus grandis.
 OS

XX US5952486-A.
 PN

XX 14-SEP-1999.
 PD

XX 21-NOV-1997; 97US-00975316.
 PF

XX 11-SEP-1996; 96US-00713000.
 PR

XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 PI

XX Grierson AM, Bloksberg LN, Havukkala I;
 PI

XX WPI; 1999-527029/44.
 DR

XX Isolated DNA sequence encoding enzymes from the lignin synthetic pathway
 PT

XX useful for generating plants with an altered lignin content.
 PT

XX Example 1; Col 37-38; 48pp; English.
 PS

XX This sequence represents a cinamoyl-CoA reductase (CCR) partial cDNA
 CC

XX from Eucalyptus grandis. This enzyme is involved in the biosynthesis of
 CC

XX lignin, an insoluble polymer which is primarily responsible for the
 CC

XX rigidity of plant stems. Lignin serves as a matrix around the
 CC

XX polysaccharide components of some plant cell walls. The higher the lignin
 CC

XX content, the more rigid the plant. Lignin also plays a role in disease
 CC

XX resistance of plants by impeding the penetration and propagation of
 CC

XX pathogenic agents. Lignin is formed by polymerisation of at least three
 CC

XX different monolignols (para-coumaryl alcohol, coniferyl alcohol and
 CC

XX sinapyl alcohol). These three monolignols are synthesised by similar
 CC

XX pathways from phenylalanine in a multistep process and are believed to be
 CC

XX polymerised into lignin via a free radical mechanism. The lignin content
 CC

XX of plants can be altered using DNA sequences encoding these enzymes.
 CC

XX Lignin content can be increased by the target plant. This could be
 CC

XX genes encoding these enzymes into the target plant. Similarly, a
 CC

XX beneficial for increasing the mechanical strength of wood. Similarly, a
 CC

XX decrease in lignin content can be obtained by transforming the target
 CC

XX plant with antisense copies of such genes. This may be beneficial in
 CC

XX plant with antisense copies of such genes. This may be beneficial in
 CC

XX trees used as forage crops for livestock (lignin is indigestible) and in
 CC

XX trees used in paper manufacture
 CC

XX Sequence 495 BP; 126 A; 121 C; 116 G; 132 T; 0 U; 0 Other;
 SQ

XX Query Match 1.0%; Score 19; DB 2; Length 495;
 XX Best Local Similarity 100.0%; Pred. No. 1.8e+02;
 XX Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 TGGTCAAGTTGGCAGG 92
 |||||
 Db 484 TGGTCAAGTTGGCAGG 466

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 03:15:12 ; Search time 7026.56 Seconds
(without alignments)
11584.370 Million cell updates/sec

Title: US-10-030-829-2

Perfect score: 1878

Sequence: 1 atgagctctagagctgctgctc.....cttcacatgaagatgatgta 1878

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl:

```

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pal:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_scs:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1878	100.0	1878	6	AX078761	AX078761 Sequence
2	1878	100.0	1909	8	BT004380	BT004380 Arabidops
3	1878	100.0	2162	8	BT002944	BT002944 Arabidops
4	963	51.3	2254	8	AF239719	AF239719 Arabidops
5	963	51.3	3275	6	AX078760	AX078760 Sequence
6	963	51.3	81365	8	AB025633	AB025633 Arabidops
7	239	12.7	650	8	ATH528171	ATH528171 Arabidops
8	22	1.2	179828	2	AC121725	AC121725 Rattus no
9	22	1.2	197909	10	AL928719	AL928719 Mouse DNA
10	22	1.2	214186	2	AC078933	AC078933 Mus muscu
11	22	1.2	218774	2	AC079182	AC079182 Mus muscu
12	22	1.2	236508	2	AC115306	AC115306 Rattus no
13	22	1.2	280016	2	AC140364	AC140364 Mus muscu
14	22	1.2	304407	2	AC073813	AC073813 Mus muscu
15	21	1.1	10225	1	AE005048	AE005048 Halobacte
16	21	1.1	50685	9	AL139004	AL139004 Human DNA
17	21	1.1	76545	2	AC080117	AC080117 Homo sapi
18	21	1.1	79554	8	NCB11823	NCB11823 Neurospor
19	21	1.1	110603	9	HS421D16	HS421D16 Human DNA
20	21	1.1	110654	2	AC146790	AC146790 Medicago
21	21	1.1	124687	9	AL592164	AL592164 Human DNA
22	21	1.1	132029	9	HS795G23	HS795G23 Homo sapi
23	21	1.1	136932	9	AC068035	AC068035 Homo sapi
24	21	1.1	143051	1	D90911	D90911 Synchocyst
25	21	1.1	160658	9	AC024258	AC024258 Homo sapi
26	21	1.1	173751	2	AL590292	AL590292 Homo sapi
27	21	1.1	175265	2	AC063964	AC063964 Homo sapi
28	21	1.1	180707	2	AC024460	AC024460 Homo sapi
29	21	1.1	183283	2	AC131075	AC131075 Mus muscu
30	21	1.1	200254	2	AC135949	AC135949 Homo sapi
31	21	1.1	202589	2	AC102629	AC102629 Mus muscu
32	21	1.1	211661	2	AC117324	AC117324 Rattus no
33	21	1.1	215773	9	AC126121	AC126121 Homo sapi
34	21	1.1	230654	2	AC115968	AC115968 Mus muscu
35	21	1.1	232071	2	AC120347	AC120347 Mus muscu
36	21	1.1	234516	2	AC108246	AC108246 Rattus no
37	21	1.1	241298	2	AC127667	AC127667 Rattus no
38	21	1.1	271898	2	AC106311	AC106311 Rattus no
39	21	1.1	300242	1	AE016790	AE016790 Pseudomon
40	20	1.1	428	8	AY135509	AY135509 Cestrum a
41	20	1.1	700	6	AX183276	AX183276 Sequence
42	20	1.1	2530	8	AK065651	AK065651 Oryza sat
43	20	1.1	4310	10	AF134918	AF134918 Mus muscu
44	20	1.1	4326	10	BC049183	BC049183 Mus muscu
45	20	1.1	4460	10	BC054532	BC054532 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS AX078761 1878 bp DNA linear PAT 22-FEB-2001
DEFINITION Sequence 2 from Patent WO0105951.
ACCESSION AX078761
VERSION AX078761.1 GI:13158380
SOURCE
ORGANISM Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
AUTHORS Beclin,C., Elmayan,T. and Vaucheret,H.
TITLE Novel sgsl plant gene and use thereof

JOURNAL Patent: WO 0105951-A 2 25-JAN-2001;
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
 AGRONOMIQUE (FR)
 FEATURES Location/Qualifiers
 source 1..1878
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"
 1..1878
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAC32419.1"
 /db_xref="GI:13158381"
 /db_xref="REMBL:CA32419"
 /translation="MSSRAGPWSKERKNTVGGYRPEVEQLVGLAETRIASSODDGEW
 EVISKNKPKGNTSGKTWVSQNSNPSPAMGQOQSGNSVGRNVSGRNGRG
 IOANSIGGRALSRKYDNNFVAPPVSRPPLGEGNMWAGSAHPTVOEPDVED
 VDNASEENDSALDDSDDLASDDYDSVSKSHSGKONKMFKEFGSLDISIEQ
 INEPRQHCPCQNGPBAIDMYNHLPIILAHARTGARVYLHRELAIVLEKIDOMRG
 ASVIPCGETYGMKGLGDEKDYETVWPVATITNTRLDKDNKMLGMGNELLEFE
 DRYEALRARHSYGPQHRGMSVLMFESSATYGLAEHLRELAEMGLRIAMGOKRSM
 FSGYROLVGFATLQODLDIPNOSHOGKTRLPFLKSYQEMVVKELROI SEBNOJLNY
 FPKLSKONKHAIVLEESLEIMSEKLRPAEDNRIVRQTKQOHNEBMDHDFE
 MSIKOIHERRADKEENPEMLQOERAKVVGQOONINPSNDCKRAEVSFIEF
 QEKMEBFYEREMLIKQEKMEBKRRHHEIFDLKEPDEALQOLMYKHLNED
 D"

ORIGIN
 Query Match 100.0%; Score 1878; DB 6; Length 1878;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGGCTGCTCCATGCTTAAAGAAAAGAAAGCTTCAGGGTGTATAGGCT 60
 1 ATGAGTTCTAGGGCTGCTCCATGCTTAAAGAAAAGAAAGCTTCAGGGTGTATAGGCT 60
 DB 61 GAGGTGAACAGTTGGTTCAGAGTTGGCAGGAGCAGACCTGGCTTCCACAAATGAT 120
 61 GAGGTGAACAGTTGGTTCAGAGTTGGCAGGAGCAGACCTGGCTTCCACAAATGAT 120
 QY 121 GGAGAGAGTGGAGGTCATTTCCAGAAAGAACAAACAAACAGAGAAACATTTCTGA 180
 121 GGAGAGAGTGGAGGTCATTTCCAGAAAGAACAAACAAACAGAGAAACATTTCTGA 180
 DB 121 GGAGAGAGTGGAGGTCATTTCCAGAAAGAACAAACAAACAGAGAAACATTTCTGA 180
 121 GGAGAGAGTGGAGGTCATTTCCAGAAAGAACAAACAAACAGAGAAACATTTCTGA 180
 QY 181 AAAAATGGGTTCTCAAAATTTGAAATCTCTTGAAGCTTGGGTTGTCAAGCAAGG 240
 181 AAAAATGGGTTCTCAAAATTTGAAATCTCTTGAAGCTTGGGTTGTCAAGCAAGG 240
 DB 181 AAAAATGGGTTCTCAAAATTTGAAATCTCTTGAAGCTTGGGTTGTCAAGCAAGG 240
 181 AAAAATGGGTTCTCAAAATTTGAAATCTCTTGAAGCTTGGGTTGTCAAGCAAGG 240
 QY 241 AGAGGTAGCAAGTATCTGGGAGAGGAAACATGTATCCGGAGAGGTAAACGCAATG 300
 241 AGAGGTAGCAAGTATCTGGGAGAGGAAACATGTATCCGGAGAGGTAAACGCAATG 300
 DB 241 AGAGGTAGCAAGTATCTGGGAGAGGAAACATGTATCCGGAGAGGTAAACGCAATG 300
 241 AGAGGTAGCAAGTATCTGGGAGAGGAAACATGTATCCGGAGAGGTAAACGCAATG 300
 QY 301 CGGGGCAATTCAGCTAACATATCTGGTGGGAGCAGACGTTGAGCAGAAAGTATGAT 360
 301 CGGGGCAATTCAGCTAACATATCTGGTGGGAGCAGACGTTGAGCAGAAAGTATGAT 360
 DB 301 CGGGGCAATTCAGCTAACATATCTGGTGGGAGCAGACGTTGAGCAGAAAGTATGAT 360
 301 CGGGGCAATTCAGCTAACATATCTGGTGGGAGCAGACGTTGAGCAGAAAGTATGAT 360
 QY 361 AACTTTGTGCAACCCCACTGTATCTCGCCCTCTTTGAGAAGAGATGAAATGGCAG 420
 361 AACTTTGTGCAACCCCACTGTATCTCGCCCTCTTTGAGAAGAGATGAAATGGCAG 420
 DB 361 AACTTTGTGCAACCCCACTGTATCTCGCCCTCTTTGAGAAGAGATGAAATGGCAG 420
 361 AACTTTGTGCAACCCCACTGTATCTCGCCCTCTTTGAGAAGAGATGAAATGGCAG 420
 QY 421 GCAAGAGAGGTTCTGCTCAGACACACAGTGTGACAGAGTTCTCTGACGTCGAGATGAT 480
 421 GCAAGAGAGGTTCTGCTCAGACACACAGTGTGACAGAGTTCTCTGACGTCGAGATGAT 480
 DB 421 GCAAGAGAGGTTCTGCTCAGACACACAGTGTGACAGAGTTCTCTGACGTCGAGATGAT 480
 421 GCAAGAGAGGTTCTGCTCAGACACACAGTGTGACAGAGTTCTCTGACGTCGAGATGAT 480
 QY 481 GTGATATATCTTCTGAGAGAGAGATGATCCGATCTTTGATGATGATGATGATGATGAT 540
 481 GTGATATATCTTCTGAGAGAGAGATGATCCGATCTTTGATGATGATGATGATGATGAT 540
 DB 481 GTGATATATCTTCTGAGAGAGAGATGATCCGATCTTTGATGATGATGATGATGATGAT 540
 481 GTGATATATCTTCTGAGAGAGAGATGATCCGATCTTTGATGATGATGATGATGATGAT 540
 QY 541 CTTCGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 541 CTTCGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 DB 541 CTTCGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 541 CTTCGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
 QY 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGATGCTTGATCGACAGATTAAT 660
 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGATGCTTGATCGACAGATTAAT 660

DB 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGATGCTTGATCGACAGATTAAT 660
 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGATGCTTGATCGACAGATTAAT 660
 QY 661 GAACCAAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 661 GAACCAAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 DB 661 GAACCAAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 661 GAACCAAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 QY 721 TATACCTGACCTCTTACTAGCTCATGCGAGGACAAAGAGGATGAGGATTAAGCTC 780
 721 TATACCTGACCTCTTACTAGCTCATGCGAGGACAAAGAGGATGAGGATTAAGCTC 780
 DB 721 TATACCTGACCTCTTACTAGCTCATGCGAGGACAAAGAGGATGAGGATTAAGCTC 780
 721 TATACCTGACCTCTTACTAGCTCATGCGAGGACAAAGAGGATGAGGATTAAGCTC 780
 QY 781 CATAGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 781 CATAGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 DB 781 CATAGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 781 CATAGAGATTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 QY 841 CTTGTGTGATGATTTATGAGGAGTGAAGGTTGGGTGAGATGAAAGATTAAGTAA 900
 841 CTTGTGTGATGATTTATGAGGAGTGAAGGTTGGGTGAGATGAAAGATTAAGTAA 900
 DB 841 CTTGTGTGATGATTTATGAGGAGTGAAGGTTGGGTGAGATGAAAGATTAAGTAA 900
 841 CTTGTGTGATGATTTATGAGGAGTGAAGGTTGGGTGAGATGAAAGATTAAGTAA 900
 QY 901 ATTGTCTGCTCCAAATGTCATCATGATGATGATGATGATGATGATGATGATGATGAT 960
 901 ATTGTCTGCTCCAAATGTCATCATGATGATGATGATGATGATGATGATGATGATGAT 960
 DB 901 ATTGTCTGCTCCAAATGTCATCATGATGATGATGATGATGATGATGATGATGATGAT 960
 901 ATTGTCTGCTCCAAATGTCATCATGATGATGATGATGATGATGATGATGATGATGAT 960
 QY 961 AAGTGTGCTGCGATGCGCAACCAAGAGCTGCTGAATTCCTGACAAAGTATGAGGCTTT 1020
 961 AAGTGTGCTGCGATGCGCAACCAAGAGCTGCTGAATTCCTGACAAAGTATGAGGCTTT 1020
 DB 961 AAGTGTGCTGCGATGCGCAACCAAGAGCTGCTGAATTCCTGACAAAGTATGAGGCTTT 1020
 961 AAGTGTGCTGCGATGCGCAACCAAGAGCTGCTGAATTCCTGACAAAGTATGAGGCTTT 1020
 QY 1021 AGAGCAGCCCAATTCCTAATGTCACAGGCGCATGCTGGAGATGAGGTTCTGATGTTGAG 1080
 1021 AGAGCAGCCCAATTCCTAATGTCACAGGCGCATGCTGGAGATGAGGTTCTGATGTTGAG 1080
 DB 1021 AGAGCAGCCCAATTCCTAATGTCACAGGCGCATGCTGGAGATGAGGTTCTGATGTTGAG 1080
 1021 AGAGCAGCCCAATTCCTAATGTCACAGGCGCATGCTGGAGATGAGGTTCTGATGTTGAG 1080
 QY 1081 AGCAGTGCACATGCTATTTGAGAGCCGACGCTCCACCGGAGTTAGCTGAGATGAGG 1140
 1081 AGCAGTGCACATGCTATTTGAGAGCCGACGCTCCACCGGAGTTAGCTGAGATGAGG 1140
 DB 1081 AGCAGTGCACATGCTATTTGAGAGCCGACGCTCCACCGGAGTTAGCTGAGATGAGG 1140
 1081 AGCAGTGCACATGCTATTTGAGAGCCGACGCTCCACCGGAGTTAGCTGAGATGAGG 1140
 QY 1141 TTAGATTAATTTGCTGGGTTGAGAGCGCAGATGATTTTCTGAGAGTGTTCGCAATG 1200
 1141 TTAGATTAATTTGCTGGGTTGAGAGCGCAGATGATTTTCTGAGAGTGTTCGCAATG 1200
 DB 1141 TTAGATTAATTTGCTGGGTTGAGAGCGCAGATGATTTTCTGAGAGTGTTCGCAATG 1200
 1141 TTAGATTAATTTGCTGGGTTGAGAGCGCAGATGATTTTCTGAGAGTGTTCGCAATG 1200
 QY 1201 TATGCTCTCTTTCAGACAGAGAGATGCTGACATTCATCAACACTCTCAAGGCAAA 1260
 1201 TATGCTCTCTTTCAGACAGAGAGATGCTGACATTCATCAACACTCTCAAGGCAAA 1260
 DB 1201 TATGCTCTCTTTCAGACAGAGAGATGCTGACATTCATCAACACTCTCAAGGCAAA 1260
 1201 TATGCTCTCTTTCAGACAGAGAGATGCTGACATTCATCAACACTCTCAAGGCAAA 1260
 QY 1261 ACAAGGCTGAATTTGAGAGTTGAATCATACCAAGAGATGTTGAAGAGCTGAGGAG 1320
 1261 ACAAGGCTGAATTTGAGAGTTGAATCATACCAAGAGATGTTGAAGAGCTGAGGAG 1320
 DB 1261 ACAAGGCTGAATTTGAGAGTTGAATCATACCAAGAGATGTTGAAGAGCTGAGGAG 1320
 1261 ACAAGGCTGAATTTGAGAGTTGAATCATACCAAGAGATGTTGAAGAGCTGAGGAG 1320
 QY 1321 ATCTGTGAGACATATGACAGCTGAATCTTTAAGAACAGCTTCAAAACAGAACAG 1380
 1321 ATCTGTGAGACATATGACAGCTGAATCTTTAAGAACAGCTTCAAAACAGAACAG 1380
 DB 1321 ATCTGTGAGACATATGACAGCTGAATCTTTAAGAACAGCTTCAAAACAGAACAG 1380
 1321 ATCTGTGAGACATATGACAGCTGAATCTTTAAGAACAGCTTCAAAACAGAACAG 1380
 QY 1381 CAGGCAAGGCTTGAAGAACTCTGGAATTTAGAGGAGAACTGGTGAAGTGA 1440
 1381 CAGGCAAGGCTTGAAGAACTCTGGAATTTAGAGGAGAACTGGTGAAGTGA 1440
 DB 1381 CAGGCAAGGCTTGAAGAACTCTGGAATTTAGAGGAGAACTGGTGAAGTGA 1440
 1381 CAGGCAAGGCTTGAAGAACTCTGGAATTTAGAGGAGAACTGGTGAAGTGA 1440
 QY 1441 GAGGATTAATCGATGATGAGACAGAACTAAGATGACAGATGAACAGAACAGGAGAG 1500
 1441 GAGGATTAATCGATGATGAGACAGAACTAAGATGACAGATGAACAGAACAGGAGAG 1500
 DB 1441 GAGGATTAATCGATGATGAGACAGAACTAAGATGACAGATGAACAGAACAGGAGAG 1500
 1441 GAGGATTAATCGATGATGAGACAGAACTAAGATGACAGATGAACAGAACAGGAGAG 1500
 QY 1501 ATGATGACACAGACAGGTTTTCATGATTTCAATCAACAGATCCTGAAAGAGAGC 1560
 1501 ATGATGACACAGACAGGTTTTCATGATTTCAATCAACAGATCCTGAAAGAGAGC 1560
 DB 1501 ATGATGACACAGACAGGTTTTCATGATTTCAATCAACAGATCCTGAAAGAGAGC 1560
 1501 ATGATGACACAGACAGGTTTTCATGATTTCAATCAACAGATCCTGAAAGAGAGC 1560
 QY 1561 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAGCTGCCAAGGTTTGGCCAG 1620
 1561 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAGCTGCCAAGGTTTGGCCAG 1620
 DB 1561 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAGCTGCCAAGGTTTGGCCAG 1620
 1561 GCAAGAGAGAGATTTGAGATGTTGACAGACAGAGAGCTGCCAAGGTTTGGCCAG 1620
 QY 1621 CAGCAGACAGACATTAATCCTCTAGCAATGACATGATGCGAAGAGAGCTGAGAGAGT 1680
 1621 CAGCAGACAGACATTAATCCTCTAGCAATGACATGATGCGAAGAGAGCTGAGAGAGT 1680
 DB 1621 CAGCAGACAGACATTAATCCTCTAGCAATGACATGATGCGAAGAGAGCTGAGAGAGT 1680
 1621 CAGCAGACAGACATTAATCCTCTAGCAATGACATGATGCGAAGAGAGCTGAGAGAGT 1680
 QY 1681 TCAAGCTTCAATGAGTTTCAAGAGAGAGATGAGAGATTTGTGAAAGAGGAGAGT 1740
 1681 TCAAGCTTCAATGAGTTTCAAGAGAGAGATGAGAGATTTGTGAAAGAGGAGAGT 1740

Db 1681 TCAGCTTCATGAGTTTTCAGAGAAAGATGAGAGAGTTTGTGAGAGAGGAGATG 1740
Qy 1741 CTGATTAAGATCAAGAGAAAGATGAAAGCATGAGAGAGGATCAGAGAGATA 1800
Db 1741 CTGATTAAGATCAAGAGAAAGATGAGAGATGAGAGAGGATCAGAGAGATA 1800
Qy 1801 TTTGATCTGAGAAAGATTTTGTAGAGCTTTGGAACAGCTCATGTACAAGATGCTT 1860
Db 1801 TTTGATCTGAGAAAGATTTTGTAGAGCTTTGGAACAGCTCATGTACAAGATGCTT 1860
Qy 1861 CACAAATGAAGATGATGA 1878
Db 1861 CACAAATGAAGATGATGA 1878
RESULT 2
LOCUS BT004380 1909 bp mRNA linear PLN 14-FEB-2003
DEFINITION Arabidopsis thaliana clone U20243 unknown protein (At5g23570) mRNA,
complete cds.
ACCESSION BT004380.1 GI:28393932
KEYWORDS FL1 cDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsids;
1 (bases 1 to 1909)
REFERENCE
AUTHORS Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,
Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
Ecker,J.R. and Theologis,A.
Arabidopsis Open Reading Frame (ORF) Clones
Unpublished
2 (bases 1 to 1909)
TITLE Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,
Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
Ecker,J.R. and Theologis,A.
Direct Submission
Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
JOURNAL The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RFLP CDNA (RFLP cDNA : RIKEN
Arabidopsis Full-Length cDNA): Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.
COMMENT
TITLE The Sak, Stanford, PGSC (SSP) Consortium members constructed and
sequenced the PUNI (ORF) clones using the RFLP CDNA: Yamada,K.,
Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
Palm,C.J., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
Ecker,J.R. and Theologis,A.
FEATURES
source Annotation based on July 2002 version of the Arabidopsis genome
submitted to Genbank.
Location/Qualifiers
1..1909
/organism="Arabidopsis thaliana"
/mol_type="mRNA"

ORIGIN
3'UTR
Query Match 100.0%; Score 1878; DB 8; Length 1909;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGAGTTCTAGAGGCTGTGCTCAATGCTTAAGAAAGAACGTTCAAGGCTTATAGGCTT 60
1 ATGAGTTCTAGAGGCTGTGCTCAATGCTTAAGAAAGAACGTTCAAGGCTTATAGGCTT 60
Db 1 ATGAGTTCTAGAGGCTGTGCTCAATGCTTAAGAAAGAACGTTCAAGGCTTATAGGCTT 60
Qy 1 ATGAGTTCTAGAGGCTGTGCTCAATGCTTAAGAAAGAACGTTCAAGGCTTATAGGCTT 60
Db 61 GAGGTTGAACATGTTGTTCAAGGTTTGGCAGGAGCAGAGCTGCTTTCACAAAGTAT 120
Qy 61 GAGGTTGAACATGTTGTTCAAGGTTTGGCAGGAGCAGAGCTGCTTTCACAAAGTAT 120
Db 61 GAGGTTGAACATGTTGTTCAAGGTTTGGCAGGAGCAGAGCTGCTTTCACAAAGTAT 120
Qy 121 GGAGGAGAGTGGAGAGTCTTTCGAAGAAAGAACCAAGAACCAAGAACCTTCTGGA 180
Db 121 GGAGGAGAGTGGAGAGTCTTTCGAAGAAAGAACCAAGAACCAAGAACCTTCTGGA 180
Qy 121 GGAGGAGAGTGGAGAGTCTTTCGAAGAAAGAACCAAGAACCAAGAACCTTCTGGA 180
Db 121 GGAGGAGAGTGGAGAGTCTTTCGAAGAAAGAACCAAGAACCAAGAACCTTCTGGA 180
Qy 181 AAAAAGGTTTCTCAGAAATTCATCTCTTCTTGAAGCTTGGGAGTGGCAGCAAGG 240
Db 181 AAAAAGGTTTCTCAGAAATTCATCTCTTCTTGAAGCTTGGGAGTGGCAGCAAGG 240
Qy 241 AAGAGTGAACAGTATCTGGAGAGAAACATGATCTCGGAGAGGTAAACGCAATGCT 300
Db 241 AAGAGTGAACAGTATCTGGAGAGAAACATGATCTCGGAGAGGTAAACGCAATGCT 300
Qy 241 AAGAGTGAACAGTATCTGGAGAGAAACATGATCTCGGAGAGGTAAACGCAATGCT 300
Db 241 AAGAGTGAACAGTATCTGGAGAGAAACATGATCTCGGAGAGGTAAACGCAATGCT 300
Qy 301 CGGGGATTCAGACTTAACATATCTGCTGGGAGCAGAGCTTGAAGCAAGATGATAAC 360
Db 301 CGGGGATTCAGACTTAACATATCTGCTGGGAGCAGAGCTTGAAGCAAGATGATAAC 360
Qy 361 AACTTTGTGGACCCCAACCTGATCTCGCTCTTGTGAAGAGAGATGATGGCAG 420
Db 361 AACTTTGTGGACCCCAACCTGATCTCGCTCTTGTGAAGAGAGATGATGGCAG 420
Qy 361 AACTTTGTGGACCCCAACCTGATCTCGCTCTTGTGAAGAGAGATGATGGCAG 420
Db 361 AACTTTGTGGACCCCAACCTGATCTCGCTCTTGTGAAGAGAGATGATGGCAG 420
Qy 421 GCAAGAGAGGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTGACCTGAGATGAT 480
Db 421 GCAAGAGAGGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTGACCTGAGATGAT 480
Qy 421 GCAAGAGAGGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTGACCTGAGATGAT 480
Db 421 GCAAGAGAGGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTGACCTGAGATGAT 480
Qy 481 GTGATTAATGCTTCTGAGAGAGATGATTCGATGCTTTGATGATGATGATGATG 540
Db 481 GTGATTAATGCTTCTGAGAGAGATGATTCGATGCTTTGATGATGATGATGATG 540
Qy 541 CTTGCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600
Db 541 CTTGCAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 600

QY 601 AATAAGTGTGAAAAAGTTCTTTGGACGCTGGAATGCTGTGCAATGAGAGATTAAT 660
 Db 601 AATAAGTGTGAAAAAGTTCTTTGGACGCTGGAATGCTGTGCAATGAGAGATTAAT 660
 QY 661 GAACCAACAGAGCAGTGGCATTTGTCCAGCTTGTCCAGAACGAGACCTGGTCCATCCATTTGG 720
 Db 661 GAACCAACAGAGCAGTGGCATTTGTCCAGCTTGTCCAGAACGAGACCTGGTCCATTTGG 720
 QY 721 TATTAACCTGACCCCTTCTACTAGCTCAGCGAGAGCAAAAAGAGCTTAGCGAGTTAACTTC 780
 Db 721 TATTAACCTGACCCCTTCTACTAGCTCAGCGAGAGCAAAAAGAGCTTAGCGAGTTAACTTC 780
 QY 781 CATTAAGATTTGGCTGAAGTTTATAGAAAGATCTACATAGATGAGAGGCGCATCTGTCATT 840
 Db 781 CATTAAGATTTGGCTGAAGTTTATAGAAAGATCTACATAGATGAGAGGCGCATCTGTCATT 840
 QY 841 CCTTGTGTGATGATTTATGAGAGAGGTTTGGTGAAGATGAAAGATTAATGA 900
 Db 841 CCTTGTGTGATGATTTATGAGAGAGGTTTGGTGAAGATGAAAGATTAATGA 900
 QY 901 ATTGTCTGCTCCCAATGCTCATCATGAACTAGACTGAGATTAAGACATTAACGAT 960
 Db 901 ATTGTCTGCTCCCAATGCTCATCATGAACTAGACTGAGATTAAGACATTAACGAT 960
 QY 961 AAGTGGCTGGAGTGGGCAACCAAGAGCTGCGAATTAATCTTGAACAAGTAAAGCTCTT 1020
 Db 961 AAGTGGCTGGAGTGGGCAACCAAGAGCTGCGAATTAATCTTGAACAAGTAAAGCTCTT 1020
 QY 1021 AAGAGCAGCCATTCATGATGTCACAGGCGCATCTGAGATGAGTTGTGATTTGAG 1080
 Db 1021 AAGAGCAGCCATTCATGATGTCACAGGCGCATCTGAGATGAGTTGTGATTTGAG 1080
 QY 1081 AAGAGTGGCACTGGCTATTGGAAGCGCGAACCCTTCCACCGGAGTTAGCTGAGTGGG 1140
 Db 1081 AAGAGTGGCACTGGCTATTGGAAGCGCGAACCCTTCCACCGGAGTTAGCTGAGTGGG 1140
 QY 1141 TTAGATAGAAATGCTGGGGTCAAGAGCGCAAGATTTTCTGAGAGTGTGGCCAACTG 1200
 Db 1141 TTAGATAGAAATGCTGGGGTCAAGAGCGCAAGATTTTCTGAGAGTGTGGCCAACTG 1200
 QY 1201 TATGCTCTCTTGAACGAGAGCAATGCTGACATATTCATCACTCTCAAGGCAAA 1260
 Db 1201 TATGCTCTCTTGAACGAGAGCAATGCTGACATATTCATCACTCTCAAGGCAAA 1260
 QY 1261 ACAAGGCTGAATTCGAGTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAGGCGAG 1320
 Db 1261 ACAAGGCTGAATTCGAGTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAGGCGAG 1320
 QY 1321 ATCTCTGAGAGCAATGAGAGCTGACCTTAAAGAACAGCTCCAAAACAGAACAG 1380
 Db 1321 ATCTCTGAGAGCAATGAGAGCTGACCTTAAAGAACAGCTCCAAAACAGAACAG 1380
 QY 1381 CAGCGCAAGGCTGTTGAGAAATCTCGAAATTTATGAGAGAGCTGGTGAAGTGA 1440
 Db 1381 CAGCGCAAGGCTGTTGAGAAATCTCGAAATTTATGAGAGAGCTGGTGAAGTGA 1440
 QY 1441 GAGGATTAATCGATCTGAGAGCAAGAACTAAGATCAGATGAACAGAACAGGAGAG 1500
 Db 1441 GAGGATTAATCGATCTGAGAGCAAGAACTAAGATCAGATGAACAGAACAGGAGAG 1500
 QY 1501 ATGAGATGACACACAGGTTTTTCAAGATTCATTAACAGATTCATGAAGAAAGAGAC 1560
 Db 1501 ATGAGATGACACACAGGTTTTTCAAGATTCATTAACAGATTCATGAAGAAAGAGAC 1560
 QY 1561 GCAAAAGAGAGATTTTCAGATGTTGACAGACAGAAAGCTCCAGGTTGTTGGCCAG 1620
 Db 1561 GCAAAAGAGAGATTTTCAGATGTTGACAGACAGAAAGCTCCAGGTTGTTGGCCAG 1620
 QY 1621 CAGACAGACAGAACTTAATCCCTCTAGCAATGACGATTCGGAAGAGCTGAGAGAGTG 1680
 Db 1621 CAGACAGACAGAACTTAATCCCTCTAGCAATGACGATTCGGAAGAGAGCTGAGAGAGTG 1680

QY 1681 TCAAGCTTCATCAGATTTCAAGAGAAAGATGAGAGATTTGTGGAAGAGAGGAGATG 1740
 Db 1681 TCAAGCTTCATCAGATTTCAAGAGAAAGATGAGAGATTTGTGGAAGAGAGGAGATG 1740
 QY 1741 CTGATTAAGATCAAGAGAAAGATGGAAGACATGAAGAAAGGCTATCAGAGAGATA 1800
 Db 1741 CTGATTAAGATCAAGAGAAAGATGGAAGACATGAAGAAAGGCTATCAGAGAGATA 1800
 QY 1801 TTTGATCTGGAAGAAATTTGATGAGGCTTTGGAACAGCTCATGATTAAGATGGCCTT 1860
 Db 1801 TTTGATCTGGAAGAAATTTGATGAGGCTTTGGAACAGCTCATGATTAAGATGGCCTT 1860
 QY 1861 CACATGAAGATGATTGA 1878
 Db 1861 CACATGAAGATGATTGA 1878

RESULT 3
 BT002944
 LOCUS
 DEFINITION
 Arabidopsis thaliana clone RAF14-93-K05 (R20243) unknown protein
 (At5g23570) mRNA, complete cds.
 ACCESSION
 BT002944
 VERSION
 BT002944.1 GI:27754622
 KEYWORDS
 FLI CDNA.
 SOURCE
 Arabidopsis thaliana (chale cress)
 ORGANISM
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 REFERENCE
 1 (bases 1 to 2162)
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C., Toriumi,M., Wong,C., Wu,H.C.,
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,
 Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
 Ecker,J.R. and Theologis,A.
 Arabidopsis Full Length cDNA Clones
 Unpublished
 2 (bases 1 to 2162)
 Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
 Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
 Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
 Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
 Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shim,P.,
 Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
 Ecker,J.R. and Theologis,A.
 Direct Submission

TITLE
 JOURNAL
 COMMENT
 Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and annotation of the RAF1 cDNAs (RAF1 cDNA: 'RIKEN
 Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
 Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
 Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAF1 cDNAs: Yamada,K., Chan,M.M.,
 Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S.,
 Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G.,
 Yuan,S., Chen,H., Cheuk,R., Jones,T., Kim,C.J., Nguyen,M.,
 Palm,C.J., Shim,P., Southwick,A., Tripp,M.G., Wu,T., Davis,R.W.,
 Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to
 this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP/PGEC)
 contributed equally to this work as PI.

Annotation based on July 2002 version of the Arabidopsis genome
 submitted to Genbank.

FEATURES
 Location/Qualifiers
 1..2162

```
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAF14-93-K05 (R20243)"
/ecotype="Columbia"
/notes="This clone is in a modified pBluescript vector2
(lambda PS) as a BamHI/XhoI insert."
1..2162
/gene="At5g23570"
1..91
/gene="At5g23570"
92..1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AA022157.1"
/db_xref="GI:27754623"
/translation="MSSRAGPMSKKNVGGYRPEVLOVLQGLAETRLASQDDGGEM
EYISKNNKPKGNTSGKTMVSONSNPPRPMGGQOQGRGNSVGRGNVSGRNGRG
IOANISGRGRALSRTDNNFVAPPPYSPREPLEGMMQARGSAQHTAVQEPDYEDD
VDNASEENDSDALDDSDDDLASDDYDSDVSQSHSGSRKQNTMFKKPFESLDSLEQ
INEPQRMWPCACQNGPQAIWYNLPLLAHARTGARVYKDLHRELAETLEKDLQMRG
ASVIPCGETYQWKGLEDEKDEYIWPMPVIIMNTRLDKDNKMLGWNQELLEYP
DYKVALRARHSYGPQGHGMSVLMFESSATGYLEARLHRELAEMGLDIAMGOKRSM
PSGGYROLVGFATKODLDIENHOSGKTRLEKPKSVQEMVYKELROI SEMOOLNY
FRNKLSKQNKAKVLEBSLEINSEKLRTRAEENRIVROKTKQHEONREEMAHDFE
MDSIKQIHERRDAKEENFEMLDQOERAKVVGQOQNIINSSNDCKRKAEVSSPLFE
OEKEMEPEFEREMLIKDEKEMEDMKRHBEIFDLEKEFDEALEOLMYKRLHNEED
D"
1970..2162
/gene="At5g23570"
2146
/gene="At5g23570"
/notes="not present in genomic sequence"

ORIGIN

Query Match 100.0%; Score 1878; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1878; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAGTTCTAGGGCTGTGTCATGCTAAGAAAAGAAAGTTCAAGGGTGTATAGGCT 60
DB 92 ATGAGTTCTAGGGCTGTGTCATGCTAAGAAAAGAAAGTTCAAGGGTGTATAGGCT 151
QY 61 GAGTTGAACAGTTGGTCAAGGTTTGCGAGGACGAGACTGCGCTTCTTCAAGATGAT 120
DB 152 GAGTTGAACAGTTGGTCAAGGTTTGCGAGGACGAGACTGCGCTTCTTCAAGATGAT 211
QY 121 GAGAGAGAGTGGAGAGTCAATTCCTCAAGAAACAAAGAACCAAGAAACCTTCTGA 180
DB 212 GAGAGAGAGTGGAGAGTCAATTCCTCAAGAAACAAAGAACCAAGAAACCTTCTGA 271
QY 181 AAAAATGCGGTTCTCAGAAATCGAATCTCTCTAGAGCTTGAGGTTGTGTCAGACGAA 240
DB 272 AAAAATGCGGTTCTCAGAAATCGAATCTCTCTAGAGCTTGAGGTTGTGTCAGACGAA 331
QY 241 AGAGGTAGCAACGTATCTGAGAGAGAAACAAATGTATCCGAGAGAGTAAACGCAATG 300
DB 332 AGAGGTAGCAACGTATCTGAGAGAGAAACAAATGTATCCGAGAGAGTAAACGCAATG 391
QY 301 CGGGGCAATTCAGCTTAACATATCTGTCGGGGACGAGCGTTGAGCAAGAAAGTATGATAC 360
DB 392 CGGGGCAATTCAGCTTAACATATCTGTCGGGGACGAGCGTTGAGCAAGAAAGTATGATAC 451
QY 361 AACTTTGTGACACCCCACTGATCTCGCCCTCTTTGGAAGAGATGATGATGACG 420
DB 452 AACTTTGTGACACCCCACTGATCTCGCCCTCTTTGGAAGAGATGATGATGACG 511
QY 421 GCAAGAGAGGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTTGAAGTGAAGTAT 480
DB 512 GCAAGAGAGGTTTGTCTCAGACACAGCTGTGACAGAGTTTCTTGAAGTGAAGTAT 571
```

```
QY 481 GTGATTAATGCTTCTGAGAGAGATGATTCCTGATGCTTTGGATGATTTCTGATGACGAC 540
DB 572 GTGATTAATGCTTCTGAGAGAGATGATTCCTGATGCTTTGGATGATTTCTGATGACGAC 631
QY 541 CTTCGAGATGATGATTAATGATCTCGAGTGAAGTCAAAAGAGCCATGATCAAGAAAGCAG 600
DB 632 CTTCGAGATGATGATTAATGATCTCGAGTGAAGTCAAAAGAGCCATGATCAAGAAAGCAG 691
QY 601 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGGATAGCTTTCATTCAGACAGATTAAT 660
DB 692 AATAAGTGTTCAAAAAGTTCTTTGGCAGCTTGGATAGCTTTCATTCAGACAGATTAAT 751
QY 661 GAACCAAGAGGACATGGGATGTCAGGCTTGTCAAGACGACCTGTGCTCATGATG 720
DB 752 GAACCAAGAGGACATGGGATGTCAGGCTTGTCAAGACGACCTGTGCTCATGATG 811
QY 721 TATTAACCTGACACCTCTACTAGCTCATGCGAGAACAAAAGAGCTAGGCGATTAACTTC 780
DB 812 TATTAACCTGACACCTCTACTAGCTCATGCGAGAACAAAAGAGCTAGGCGATTAACTTC 871
QY 781 CATAGAGATTTGGCTGAAGTTTATAGAAAAGATCTACAGATGAGAGCGCATCTGCTATT 840
DB 872 CATAGAGATTTGGCTGAAGTTTATAGAAAAGATCTACAGATGAGAGCGCATCTGCTATT 931
QY 841 CCTTGCGGTGAGATTATATGAGGACGTGAAGGGTTTGGGTGAGATGAAGAAAGATTATGAA 900
DB 932 CCTTGCGGTGAGATTATATGAGGACGTGAAGGGTTTGGGTGAGATGAAGAAAGATTATGAA 991
QY 901 ATTGTCGTGCTCCATGTCATCATGATTAAGTGAATGATGATTAAGACATTAACGAT 960
DB 992 ATTGTCGTGCTCCATGTCATCATGATTAAGTGAATGATGATTAAGACATTAACGAT 1051
QY 961 AAGTGGCTTGGCATGAGGCAACCAAGAGCTGTGGAATACTTGAACAAGTATGAGGCTCTT 1020
DB 1052 AAGTGGCTTGGCATGAGGCAACCAAGAGCTGTGGAATACTTGAACAAGTATGAGGCTCTT 1111
QY 1021 AGAGCAGGCAATTCCTAATGTCACACAGGCGCATGCGGAGATGATGTTCTGATGTTGAG 1080
DB 1112 AGAGCAGGCAATTCCTAATGTCACACAGGCGCATGCGGAGATGATGTTCTGATGTTGAG 1171
QY 1081 AGCAGTGCACATGCGTATTTGAGAGGCGCAACGCTTCCACCGGAGTTAGCTGATGAGG 1140
DB 1172 AGCAGTGCACATGCGTATTTGAGAGGCGCAACGCTTCCACCGGAGTTAGCTGATGAGG 1231
QY 1141 TTAGATGAATTTGCTGGGCTCAAGAGGCAAGTATGTTTCTGAGAGTGTGGCCAACTG 1200
DB 1232 TTAGATGAATTTGCTGGGCTCAAGAGGCAAGTATGTTTCTGAGAGTGTGGCCAACTG 1291
QY 1201 TATGCTTCCTTGAACGAGAGATGTCGACATTAATCAATCAACATCTCAAGGCAAA 1260
DB 1292 TATGCTTCCTTGAACGAGAGATGTCGACATTAATCAATCAACATCTCAAGGCAAA 1351
QY 1261 ACAAGGCTGAATTCGATTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAGGCGAG 1320
DB 1352 ACAAGGCTGAATTCGATTTGAATCATACCAAGAGATGTTGTAAAGAGCTGAGGCGAG 1411
QY 1321 ATCTGTAGGACATTCAGACGCTGAATCTTTAAGAACACAGCTTCAAAACAGAACAG 1380
DB 1412 ATCTGTAGGACATTCAGACGCTGAATCTTTAAGAACACAGCTTCAAAACAGAACAG 1471
QY 1381 CAGGCAAGTGTCTTGAAGAACTCTCGAATAATTAAGAGGAGAACTGCTGTGAACCTGCA 1440
DB 1472 CAGGCAAGTGTCTTGAAGAACTCTCGAATAATTAAGAGGAGAACTGCTGTGAACCTGCA 1531
QY 1441 GAGATTAATTCGATTCGTGAGACAGAGAACTTAAGATGACATGAACAGAAACGAGAAAG 1500
DB 1532 GAGATTAATTCGATTCGTGAGACAGAGAACTTAAGATGACATGAACAGAAACGAGAAAG 1591
QY 1501 ATGATGCAACACAGAGGTTTTTCAATGATTAATCAACAGATCAATGAAGAAAGAGAC 1560
DB 1592 ATGATGCAACACAGAGGTTTTTCAATGATTAATCAACAGATCAATGAAGAAAGAGAC 1651
```


GENSCAN (Chris Burge, MIT, <http://CCR-081.mit.edu/GENSCAN.html>),
Necenz2 (S.M. Hebsgaard, et al., CBS, Technical University of
Denmark, <http://www.cbs.dtu.dk/services/Necenz2/>) and
SplicePredictor (Volker Brendel, Stanford University,
<http://genome1.zool.iastate.edu/cgi-bin/sp.cgi>).
Genes encoding tRNAs are predicted by tRNAscan-SE
(Sean Eddy, Washington University School of Medicine, St. Louis,
<http://genome.wustl.edu/eddy/tRNAscan-SE/>).
This sequence may not be the entire insert of this clone. It may be
shorter because we remove overlaps between neighboring submissions.
The 5' clone is K19M13 and the 3' clone is M9011.

FEATURES
source

```
1..81365
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="5"
/clone_1ib="M1c5ui P1"
join(3184..3403,3474..3595,3681..3723,3777..3898,
4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)
/note="gb|AA01580.1
gene_id:M0M1.24
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97251.1"
/db_xref="GI:8809710"
/translation="MHVQGTSDVDSPAHKVDPEPPKHVQLEPISTKVDDAQNPS
YGSADDPSSNSPLISVFEPPSSFFSEGQDDPFGIEDQISGEPPGHELOACG
YSINGTSCNFEMVCHLEDGVSNIYIDESSERTSNAAGAPVYLVADVDVYLAEV
QGLDNRNRKGRFARFGKYVDLVLDVHTLTFSDCFSEPFMMFAICYEELISG
ACEDDMOSNIETKTHGTGASVKSVLAVFVIMWRAITISIRBGYSIKCISDITAE
KKSASTVTIPFGQPAELVITISDSGSHSLRDNSSPDIGSRDELVTLRIFIKRW
FKALNSFYHFCSCSKSRGNL"
join(6317..6428,6508..6629,6714..6756,6841..6932,
7121..7169,7336..7466,7550..7663,7785..7882,8112..8181)
/note="gb|AA01580.1
gene_id:M0M1.23
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97250.1"
/db_xref="GI:8809709"
/translation="MQPQSYSPALNADDETPISKTAVNDTKNIEYFDDGAPL
PALBNQISGEPPGHELOACGYSINGTSCNFEMVCHLEDGVSNIYIDAKKPNLTAVT
ADVDGICLAEVQPLDNRKGLVFPANDRKIAKCHPEMOSNIDKTLHTGASVKG
SLAIGFVHMEATLISIEREGYTIKCNNDLTITKFSASTAVKIPPEKPAELVITISD
GSEHCIRVNEWDIRLCKEERVLALVYVVFQKL"
complement(join(8642..8705,8794..8863,8978..9387,
9584..10347))
/note="gene_id:M0M1.22
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97249.1"
/db_xref="GI:8809708"
/translation="MSWAKGSSGWTAPLQKQKGLSEVBGDPFPVSVTSVNASF
GVGRGLRNHPESEKSFSSVLLPSPRFPAITENKOCNGERGCGCKRCKDTLSLPNS
HDLAFTKLEMSWADNDLIRVLLSTEDFEMALAFIKGMVSGSGKDEDEPSKIEGY
SDNRSSEYREFEKTIVTSVKMAARSTFEDAGKYDLSGSSFLVNASDNKFPDI
SELDSTLQRLQSPTEPEWEDDLVSHRKDALKVMRSASNSRAQNAFGYVDHSA
KQHSDEAREDMLAAELNBAKAKITGINKNDITKLDLHGLHATEVQALQERLQW
IGHFTVNSVSPNRSRKNALRSASQSPFGDLDEBGNHCORTSRELRNSLQVITG
IKHSRQASLPLAAYTFEDNRKRFDETRPGVITVRPFRHS"
complement(12417..13424)
/note="contains similarity to unknown protein
gb|AA027018.1
gene_id:M0M1.21"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97248.1"
```

```
/db_xref="GI:8809707"
/translation="MATDSQPNQKLTLEKTRIALTVISTMTWNAQRPDGTINREFLR
LEDFRAPNPKNVINSVSDVFOQSRDLNRLTLPVSGKIPVVFHGAGFAPLS
PNAYPVNDCRRPARKLPAVYISVYRLAPVOYDDGDFDLKYLENHGSLTPA
NADLSRCFPADSAGNIAHVAIRICGEPSSFPAYVLIGLISGPPGGEETAE
KQLVGAPLVSFDRDTCWAKLNLDRAVAVGGBNAYDISGLDPEPMVYVAGDPL
KDWORSYEWMLCKGKATLLIEYPMFAFYIFPELPAAGOLIMIKIKOFVERVASLS
A"
join(14922..14981,15072..15167,15563..15666,15755..15891,
16096..16178)
/note="gene_id:M0M1.20"
/codon_start=1
/evidence=not experimental
/product="50S ribosomal protein L24"
/protein_id="BAA97247.1"
/db_xref="GI:8809706"
/translation="WGRKAAEKLIIRHKILRGDNVMTIRGDKSETGTXAVIRSONR
VIVEGKLIKHKIRGPDHGGIFTEVPELHASNQVQVDPVGRPKGVKRYLEDGK
VRVARGTSGSIIIPREILIKRATPREPTGPKDTPMEFWEGYDAKTGKMPDL"
join(16823..16952,17143..17320,17412..17493,17675..17851,
17943..18110,18199..18390)
/note="gene_id:M0M1.19"
/codon_start=1
/evidence=not experimental
/product="26S proteasome, non-ATPase regulatory subunit"
/protein_id="BAA97246.1"
/db_xref="GI:8809705"
/translation="MERLQRIFGAGGGLGHASPDSPITLDSBOVYSSIALIKMLKHG
RAGVMEYMGMLGERVDYRIVRYVDVYAMQSGVGSVEAYDHYFQNMIDMLKQTS
RPEVAVGMYHSHPGFCMLSGVDINTQOSFALNQAARAVAVDPIQSFKGVADIAFR
SINPTIMLQGEPRQTSNIGLHINKPSIQALIHGNRYISIAINRYRNELEEKMLN
LHKKMTDGLTRPFDTSKTNQEQVQMLSLAKYKNAVQVEBELSPEKLAIVVGR
QDAKHLIEHNSNIMSNIYOTLGTMLTVVF"
complement(join(18965..19043,19240..19298,19419..19536,
19666..19739,19908..19961,20387..20530))
/note="contains similarity to unknown protein
emb|CAA2897.1
gene_id:M0M1.18"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97245.1"
/db_xref="GI:8809704"
/translation="MDKNQOAIKKKMLVMEVEDEQQADESSLSFMEDLNRCAL
TTKQRFYFAICLSAGLTCTLLSMLVFNPAKFGITFTLGMALGSAFPLGPOROY
TMMDLPARIYATVATYIASIITLFCALYVRNKLTLTALITLEFTGLVYSYIPFAR
TMVSKIMTCCDTER"
join(22672..22634,23708..23999,24111..24355,24458..24624,
24715..24925)
/note="emb|CAB2356.1
gene_id:M0M1.17
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97244.1"
/db_xref="GI:8809703"
/translation="MSSRAGPMSKEKNVQGYRPEVQILVOGLAGTRIASQDDGGEW
EVISKKNKPKGNTGKTWVSNSNPAPMGQOQGRGNSGRANYSGRNGNGRG
IOANTSGRGRLSRKYDNNPVAAPPVASPLEGNNWARGGSAHTVQEPVEDD
VDNASSEENDSDALDDSDDLASDDYSDVQSKSGSKQNKPKFKFGSIDSLSIBQ
INEPQONHCPCQKQPGALIDWYNLHPLIARTRIGARVKLHRELAVLEKDLQMRG
ASVLPQGIYVYVGPQGHMGVLMESASPTVLEKRLHRELAEWGLRIAMQGRSM
DKVEALRARSYVGPQGHMGVLMESASPTVLEKRLHRELAEWGLRIAMQGRSM
FGGVRQVLYGLFATKQDDIDFNQSGSTRKFLFKLSQEWVVKELROI SEDNOOLNY
FKNTLSKQNKAKYALBESLEIMSEKLTARTADNTRVORTWQCHQONREMDAHBF
MDSTIQIRERDPAKENFEMIQQQRARVVGQOQONINPSSNDCKRAEVSSTIEF
QKEMEEFVEEREMLEIKQEKMEKMRKRRHBEIFDLKEPEDEALEQMLRYVGLNHED
D"
join(25614..25838,26061..26207,26324..26438,26516..26610,
26710..26813,27532..27765,27943..28036,28115..28305,
28385..28604,28688..28855,28944..29092)
/note="gb|AAC97420.1
gene_id:M0M1.16
strong similarity to unknown protein"
```



```

/codon_start=1
/evidence=not experimental
/protein_id="BAA97243.1"
/db_xref="GI:8809702"
/translation="MAPPAQGTAAVAEAAADGAPPOQGRGSGTISGIVRAVWYF
ASKFSPKQKMPDPAFQMLTNLPHKESLDWMFYLBQEFNFQNDRALYMEHTN
IPIVAWTPESIRTKSLIYVPSSETLQNNGLYAHIFAPSGSDIDTDEYQWISFNR
THAVATYPPKQKSKSLGSDKSDSEPEVKYDKSPKSEPEVEMISLKPEN
VTINLVDDPTQADLVEPTTGNYYPTIYFNEFMILRDKFIPIVNETVSEIPLNTEISF
ISMWKQI.FQOVQDSFQOMRYSYGMIDSESDLKVEPLEGNYXLLGITMFMHSVF
DPLAFKNDIQFNKSKSMEGLSAKSVLNFQCFITFLYLDNDTSMILASSGVIC
IEFWKIGRAMRLEVDSEMIPELRPHDEBSVANSNTKEVDIAIFLSVLLLVIGL
SIYSLAVRHRKSMYSMILSLTSVYMGFIMCQPLFNLYDLKSVNHLPMQMYKFK
LNTIIDLFFAFVYIKMPLILRLSVFRDVI.FLYLYRWVYPPDKTRVIEFGGDEET
AEKLLITEKEEDDKXTN"
complement(join(29439..29576,29661..29885,29985..30268,
30341..30493,30579..30722,30956..31484))
/note="gene_id:MOM1.15

```

Query Match 51.3%; Score 963; DB 8; Length 81365;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

1 ATGAGTTCTAGGGCTGTGTCATGTCCTAAGAAAAGACGTTCAAGGTTATAGCCT 60
22672 ATGAGTTCTAGGGCTGTGTCATGTCCTAAGAAAAGACGTTCAAGGTTATAGCCT 22731
61 GAGGTTGAACAGTTGGTTCAAGGTTGGAGGAGCAGACCTGGCTTTTCAACAATGAT 120
22732 GAGGTTGAACAGTTGGTTCAAGGTTGGAGGAGCAGACCTGGCTTTTCAACAATGAT 22791
121 GGAGGAGAGTGGAGGATCATTTCCAAAGAAACAAAGAACCAACCAAGGAAACCTTCTGGA 180
22792 GGAGGAGAGTGGAGGATCATTTCCAAAGAAACAAAGAACCAACCAAGGAAACCTTCTGGA 22851
181 AAAAATGGGTTTCTCAGAAATTCGATCCTCTAGAGCTTGGGTTGTCAGACGAGCAAGG 240
22852 AAAAATGGGTTTCTCAGAAATTCGATCCTCTAGAGCTTGGGTTGTCAGACGAGCAAGG 22911
241 AGAGGTACGACGATTCCTGGAGAGAGAAACATGATCCGGAGAGGTATACGGCAATGCT 300
22912 AGAGGTACGACGATTCCTGGAGAGAGAAACATGATCCGGAGAGGTATACGGCAATGCT 22971
301 CGGGGCAATTCAGGATCAATATCTGGTGGGAGCAGAGGTTGAGAGAAAGATATATAC 360
22972 CGGGGCAATTCAGGATCAATATCTGGTGGGAGCAGAGGTTGAGAGAAAGATATATAC 23031
361 AACTTTGTGACACCCCACTGTATCTGCCCTCTTGGAGAGAGATGATTTGGCAG 420
23032 AACTTTGTGACACCCCACTGTATCTGCCCTCTTGGAGAGAGATGATTTGGCAG 23091
421 GCAAGAGAGAGTTTCTGCTCAGACACAGCTGTGCAAGAGTTTCTGACCTGAGATGAT 480
23092 GCAAGAGAGAGTTTCTGCTCAGACACAGCTGTGCAAGAGTTTCTGACCTGAGATGAT 23151
481 GTGATATAGCTTCTGAGGAGAGAGATGATTCGATGCTTGGATGATTCGATGACGAC 540
23152 GTGATATAGCTTCTGAGGAGAGAGATGATTCGATGCTTGGATGATTCGATGACGAC 23211
541 CTTGCAAGTGAATGATATGACTCGATGATGATCAAAAGAGCATGATCAAGAAAGCAG 600
23212 CTTGCAAGTGAATGATATGACTCGATGATGATCAAAAGAGCATGATCAAGAAAGCAG 23271
601 AATAAGTGTTCACAAAAGTTCTTTGGCAGCTTGGATGATGCTTCTGATGAGCAGATTAAT 660
23272 AATAAGTGTTCACAAAAGTTCTTTGGCAGCTTGGATGATGCTTCTGATGAGCAGATTAAT 23331
661 GAACACAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 720
23332 GAACACAGAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 23391
721 TATAACTGTCACCTCTACTAGCTGATGAGAGCAAAAGAGCTATAGGCGAGTAACTC 780

```

```

Db 23392 TATAACTGTCACCTCTACTAGCTGATGAGAGCAAAAGAGCTATAGGCGAGTAACTC 23451
Qy 781 CATAGAGATTTGGCTGAAGTTTATAGAAAAGATCTACAGATAGAGGCGCATCTGTCATT 840
Db 23452 CATAGAGATTTGGCTGAAGTTTATAGAAAAGATCTACAGATAGAGGCGCATCTGTCATT 23511
Qy 841 CCTTGTGTGATGATTTATGAGGAGCAGTGAAGGCTTTGGTGAAGATGAAAAGATTATGAA 900
Db 23512 CCTTGTGTGATGATTTATGAGGAGCAGTGAAGGCTTTGGTGAAGATGAAAAGATTATGAA 23571
Qy 901 ATTGTCTGGCTCCCAATGTCATCATGATGATCTAGATGATGATGATGATGATGATGAT 960
Db 23572 ATTGTCTGGCTCCCAATGTCATCATGATGATGATGATGATGATGATGATGATGATGAT 23631
Qy 961 AAG 963
Db 23632 AAG 23634

```

RESULT 7
 ATHS28171 650 bp DNA linear PIN 29-MAR-2003
 LOCUS
 DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone 157C08.

ACCESSION AJ528171 GI:26796431
 VERSION AJ528171.1
 KEYWORDS left border; T-DNA flanking sequence.
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
 Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Sanson, F., Chauvin, S., Bechold, N., Cruaud, C., Dekose, R., Pelletier, G., Lepoint, L., Caboche, M. and Lecharny, A.
 T-DNA integration into the Arabidopsis genome depends on sequences of pre-integration sites
 EMBO Rep. 3 (12), 1152-1157 (2002)
 MEDLINE 22363535
 PUBMED 12446565
 JOURNAL 2 (bases 1 to 650)
 REFERENCE Balzergue, S.
 TITLE Direct Submission

AUTHORS Submitted (21-NOV-2002) Balzergue S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
 JOURNAL PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment (s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).
 FEATURES location=info.infobiogen.fr.

FEATURES
 source 1..650
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /cultivar="Wassiljewskij3"
 /db_xref="taxon:3702"
 /clone="157C08"
 /clone_1lb="Arabidopsis thaliana T-DNA insertion lines"
 misc_feature 1..650
 /note="T-DNA flanking sequence
 left border"

ORIGIN
 Query Match 12.7%; Score 239; DB 8; Length 650;
 Best Local Similarity 99.7%; Pred. No. 9.9e-133;
 Matches 289; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY	962	AGTGGCTCGGATGGGCAACAAGAGCTGCTGGAATTACTTGACAAAGTATGAGGCTCTTA	1021
Db	303	AGTGGCTCGGATGGGCAACAAGAGCTGCTGGAATTACTTGACAAAGTATGAGGCTCTTA	362
QY	1022	GAGCAGCCCAATCTCTATGGTCCACAGGGGCATCTGGGGATGAGTGTTCATGTTTGA	1081
Db	363	GAGCAGCCCAATCTCTATGGTCCACAGGGGCATCTGGGGATGAGTGTTCATGTTTGA	422
QY	1142	TAGATAGAATTGCTTGGGGGTCGAAAGCCGACATGATGTTTTCGAGGTGTCCCACTGT	1201
Db	483	TAGATAGAATTGCTTGGGGGTCGAAAGCCGACATGATGTTTTCGAGGTGTTCGCAACTGT	542
QY	1202	ATGGCTTCTCTGCACGAAAGCAAGATCTGGACATATTCACATCAACACT	1251
Db	543	ATGGCTTCTCTGCACGAAAGCAAGATCTGGACATATTCACATCAACACT	592
RESULT 8	AC121725/c	AC121725	178928 bp DNA linear HTG 11-OCT-2007
LOCUS		Rattus norvegicus clone CH230-32812, *** SEQUENCING IN PROGRESS	
DEFINITION		***, 3 unordered pieces.	
ACCESSION		AC121725 GI:23804926	
VERSION		HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.	
KEYWORDS		Rattus norvegicus	
SOURCE		Rattus norvegicus	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE		1 (bases 1 to 178928)	
AUTHORS		Muzny,D.Matie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiulano,D., Anylepech,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Bunay,C., Burch,P., Butrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M., L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deremo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duvall,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Frazer,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeregs,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Haylak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huijck,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolyet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Lervan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Louieged,H., Lozado,R.J., Lu,X., Ma,D., Lorensuella,L., Louieged,H., Lozado,R.J., Lu,X., Ma,D., Maheshwari,M., Mahindaratne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mamliney,S., McLeod,M.P., McNeill,T.Z., Meenan,E., Milosavljevic,A., Miner,G., Mijic,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Nandiasa,M., Murphy,M., Nait,L., Nankervils,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwalekeme,O., Okunolu,G., Olarinuogun,A., Pal,S., Parks,K., Pasterkamp,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C., Plapper,F., Polndexter,A., Popovic,D., Prims,E., Pu,L.-L., Puazo,M., Quiriz,J., Rachlin,E., Reeves,K., Regier,M.A., Reish,R., Reilly,B., Reilly,T., Ren,Y., Reuter,R., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,	

[illegible]

```

misc_feature      /mol_type="genomic DNA"
                  /db_xref="taxon:10116"
                  /clone="CH230-32812"
                  1..1169
                  /note="wgs_end_extension
                  clone_end:Sp6"
misc_feature      27453..29430
                  /note="wgs_end_extension
                  clone_end:Sp6"
misc_feature      77608..79432
                  /note="wgs_end_extension
                  clone_end:Sp6"
misc_feature      99609..100719
                  /note="wgs_end_extension
                  clone_end:Sp6"
misc_feature      complement(118199..119034)
                  /note="clone_boundary
                  clone_end:Sp6
                  site:MboI
                  end_sequence:RXAON49TV"
misc_feature      125929..127858
                  /note="wgs_contig"
misc_feature      156410..157915
                  /note="wgs_contig"
misc_feature      complement(172023..172895)
                  /note="clone_boundary
                  clone_end:T7
                  site:MboI
                  end_sequence:RXAON49TV"
misc_feature      173824..175436
                  /note="wgs_end_extension
                  clone_end:T7"

```

ORIGIN

```

Query Match      1.2%; Score 22; DB 2; Length 178928;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 492 TTCTGAGGAGAGATGATTC 513
Db 112510 TTCTGAGGAGAGATGATTC 112489

```

```

RESULT 9
LOCUS AL928719 197909 bp DNA linear ROD 28-JAN-2003
DEFINITION Mouse DNA sequence from clone RP23-419G21 on chromosome 2, complete
sequence.
ACCESSION AL928719
VERSION AL928719.6 GI:24527684
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 197909)
REFERENCE 1
AUTHORS Leongamornlert,D.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-2003) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
On Nov 3, 2003 this sequence version replaced gi:24395353.
Sequence from the Mouse Genome Sequencing Consortium whole genome
shotgun may have been used to confirm this sequence. Sequence data
from the whole genome shotgun alone has only been used where it has
a phred quality of at least 30.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
-----

```

During sequence assembly data is compared from overlapping clones.

Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the rare assembly was confirmed by restriction digest, except on the occasion of the clone being a YAC. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TREMBL; Mp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep RP23-419G21 is from the RPII-23 Mouse BAC Library constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6

FEATURES

```

source
Location/Qualifiers
1..197909
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-419G21"
/clone_1fb="RPII-23"

```

ORIGIN

```

Query Match      1.2%; Score 22; DB 10; Length 197909;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 567 TGTGAGTCAAAAGAGCCATGGA 588
Db 34650 TGTGAGTCAAAAGAGCCATGGA 34629

```

```

RESULT 10
LOCUS AC078933 214186 bp DNA linear HTG 11-AUG-2000
DEFINITION Mus musculus chromosome 5 clone RP23-201B13 strain C57BL6/J,
WORKING DRAFT SEQUENCE, 17 unordered pieces.
ACCESSION AC078933
VERSION AC078933.1 GI:9795569
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214186)
REFERENCE 1
AUTHORS Beckettrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
Huang,M.C., Idol,J., Lee-Jin,S.-Q., Maduro,O.L., Maduro,V.B.,
Maestrian,S.D., McCloskey,J.C., McDowell,J., Ojodu,M.A., Pearson,R.,
Stantitrop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,
Tingerson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,
Welchery,K.D., and Green,E.D.
TITLE NISC Mouse Sequencing Initiative
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 214186)
AUTHORS Green,E.D.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2000) NIH Intramural Sequencing Center, 8717
Groveomnt Circle, Gaithersburg, MD 20877, USA
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: http://www.nisc.nih.gov
Contact: nisc_mouse@nhgri.nih.gov
-----

```

	misc_feature	13188..20448	/note="assembly_fragment"
	misc_feature	20549..26659	/note="assembly_fragment"
	misc_feature	26760..35349	/note="assembly_fragment"
	misc_feature	35450..43812	/note="assembly_fragment"
	misc_feature	43913..52486	/note="assembly_fragment"
	misc_feature	52587..62100	/note="assembly_fragment"
	misc_feature	62201..77345	/note="assembly_fragment"
	misc_feature	77446..87428	/note="assembly_fragment"
	misc_feature	clone_end:T7	/note="assembly_fragment"
	misc_feature	vector_side:right	/note="assembly_fragment"
	misc_feature	87529..100606	/note="assembly_fragment"
	misc_feature	100707..117904	/note="assembly_fragment"
	misc_feature	118005..135715	/note="assembly_fragment"
	misc_feature	clone_end:SP6	/note="assembly_fragment"
	misc_feature	vector_side:right	/note="assembly_fragment"
	misc_feature	135816..153408	/note="assembly_fragment"
	misc_feature	153509..181034	/note="assembly_fragment"
	misc_feature	181135..214186	/note="assembly_fragment"
ORIGIN			
Query Match	1.2%; Score 22; DB 2; Length 214186;		
Best Local Similarity	100.0%; Pred. No. 2.1;		
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0.			
Cy	567 TGTGAGTCAAAAGAGCCATGGA 588		
Db	99680 TGTGAGTCAAAAGAGCCATGGA 99701		
RESULT 11			
LOCUS ACO79182/c	218774 bp DNA linear HTG-23-AUG-2000		
DEFINITION Mus musculus chromosome 5 clone RP23-203F6 strain C57BL6/J, WORKING			
ACCESSION DRAFT SEQUENCE, 16 unordered pieces.			
VERSION ACO79182			
KEYWORDS ACO79182.1 GI:9886001			
SOURCE HTG; HTGS_PHASE1; HTGS_DRAFT.			
ORGANISM Mus musculus (house mouse)			
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
	1 (bases 1 to 218774)		
	Beckstrom-Sternberg,S.M., Benjamin,B., Blakeley,R.W.,		
	Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,U., Ho,S.-L.,		
	Idd,J., Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L.,		
	Maduro,V.B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,		
	Pearson,R., Stantirpop,S., Summers,T.J., Thomas,J.W., Thomas,P.J.,		
	Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L., Walker,M.A.,		
	Weberthy,K.D. and Green,E.D.		
JOURNAL NISC Mouse Sequencing Initiative			
REFERENCE Unpublished			
AUTHORS 2 (bases 1 to 218774)			
JOURNAL Direct Submission			
TITLE Submitted (23-AUG-2000) NIH Intramural Sequencing Center, 8717			
JOURNAL Grovmont Circle, Gaithersburg, MD 20877, USA			
COMMENT ----- Genome Center			
	Center: NIH Intramural Sequencing Center		
	Center code: NISC		

```

Web site: http://www.nisc.nih.gov
Contact: nisc.mouse@nih.gov
-----
Project Information
Center project name: ve
Center clone name: 203R06
-----
Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 208040 bases at least Q40
Consensus quality: 211445 bases at least Q20
Consensus quality: 213463 bases at least Q20
Insert size: 227000; agarose-fp
Insert size: 242000; pulse-field-gel
Insert size: 217274; sum-of-ctnigs
Quality coverage: 4.70x in Q20 bases; agarose-fp
Quality coverage: 4.40x in Q20 bases; pulse-field-gel
Quality coverage: 4.91x in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 16 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1 2001: contig of 2001 bp in length
2 2002: contig of 2001 bp in length
3 2101: gap of unknown length
4 2102: contig of 2769 bp in length
5 4871: gap of unknown length
6 4971: contig of 3499 bp in length
7 8470: gap of unknown length
8 8570: contig of 5663 bp in length
9 14233: gap of unknown length
10 14332: gap of unknown length
11 14333: contig of 7299 bp in length
12 21631: gap of unknown length
13 21731: gap of unknown length
14 21732: contig of 7930 bp in length
15 29661: gap of unknown length
16 29662: contig of 8381 bp in length
17 38142: gap of unknown length
18 38143: contig of 9488 bp in length
19 38243: gap of unknown length
20 47730: gap of unknown length
21 47731: contig of 8114 bp in length
22 47831: contig of 8114 bp in length
23 55944: gap of unknown length
24 55945: gap of unknown length
25 56044: gap of unknown length
26 56045: gap of unknown length
27 70548: gap of unknown length
28 70549: gap of unknown length
29 70649: contig of 14867 bp in length
30 85615: gap of unknown length
31 85616: gap of unknown length
32 103358: contig of 17743 bp in length
33 103359: gap of unknown length
34 103459: contig of 21626 bp in length
35 125084: gap of unknown length
36 125085: contig of 28172 bp in length
37 125185: gap of unknown length
38 153356: gap of unknown length
39 153357: contig of 3189 bp in length
40 153457: gap of unknown length
41 184845: gap of unknown length
42 184846: contig of 33829 bp in length
43 218774: contig of 33829 bp in length
-----
Location/Qualifiers
1. 218774
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/chromosome="5"
/clone_lib="RPCI mouse BAC library 23"
1. 2001
/note="assembly_fragment"
2102. 4870
/note="assembly_fragment"
4971. 8469
/note="assembly_fragment"
/note="assembly_fragment"

```

```

misc_feature 8570..14232
/note="assembly_fragment"
misc_feature 14333..21631
/note="assembly_fragment"
misc_feature 21732..29661
/note="assembly_fragment"
misc_feature 29762..38142
/note="assembly_fragment"
misc_feature 38243..47730
/note="assembly_fragment"
misc_feature 47831..55944
/note="assembly_fragment"
misc_feature 56045..70548
/note="assembly_fragment"
misc_feature 70649..85615
/note="assembly_fragment"
misc_feature 85616..103358
/note="assembly_fragment"
misc_feature 103459..125084
/note="assembly_fragment"
misc_feature 125085..153356
/note="assembly_fragment"
misc_feature 153357..184845
/note="assembly_fragment"
misc_feature 184846..218774
/note="assembly_fragment"
-----
ORIGIN
Query Match 1.2%; Score 22; DB 2; Length 218774;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 567 TGTGAGTCAAAAGAGCCATGCA 568
|||||
Db 76509 TGTGAGTCAAAAGAGCCATGCA 76488
-----
RESULT 12
AC115306/c 236508 bp DNA linear HTG 23-NOV-2002
LOCUS Ratrus norvegicus clone CH230-11F1, WORKING DRAFT SEQUENCE, 3
DEFINITION Rattus norvegicus (Norway rat)
UNORDERED PIECES.
ACCESSION AC115306
VERSION 4 GI:25188797
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 236508)
Muzny,D.,Marie,,Metker,M.,Lee,,Abramzon,S.,Adams,C.,Alder,J.,
Allen,C.,Allen,H.,Albrooks,S.,Amin,A.,Angiano,D.,
Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,B.,Baden,H.,
Baldwin,D.,Banderanaike,D.,Barber,M.,Barnstead,M.,Benahmed,F.,
Biswal,K.,Blair,J.,Blankenburg,K.,Blyth,K.,Calderton,M.,
Carmichael,N.,Carter,K.,Cavazos,I.,Cesar,H.,Center,A.,
Cardenas,V.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,
Davis,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dederich,D.,
Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K.,
Draeger,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,
Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Fallis,T.,Fan,G.,
Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P.,
Fraser,C.M.,Gabisi,A.,Ganta,R.,Garcia,A.,Garner,T.,Garza,M.,
Gebregorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,
Gunnarsson,P.,Haaland,W.,Hamill,C.,Hamilton,C.,Hamilton,K.,
Harvey,Y.,Havlak,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,
Hollins,B.,Howells,S.,Huylk,S.,Hume,J.,Idlebird,D.,Jackson,A.,

```

```

Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kover, C.,
Kovacs, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshaw, L., Louised, H., Lozano, R., Lu, X., Ma, J.,
Manshewari, M., Mahindartne, M., Mahmood, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Muniada, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, G., Olampunagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Shen, H.,
Sanders, M., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajic, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soza, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umami, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, D., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Weinhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstein, G., and Gibbs, R.A.

Unpublished
Direct Submission
2 (bases 1 to 236508)
Worley, K.C.
Direct Submission
Submitted (17-MAR-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 236508)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (23-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 23, 2002 this sequence version replaced gi:22772519.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information -----
Center project name: GBS
Center clone name: CH230-11F1

----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 219730 bases at least Q40
Consensus quality: 22791 bases at least Q30
Consensus quality: 22433 bases at least Q20
Estimated insert size: 225524; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

```

```

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 125318: contig of 125318 bp in length
* 125319 125418: gap of unknown length
* 125419 180275: contig of 54857 bp in length
* 180276 236508: gap of unknown length
* 180376 236508: contig of 56133 bp in length.
* Location/Qualifiers
  1. 236508 "Rattus norvegicus"
    /organism="Rattus norvegicus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10116"
    /clone="CH230-11F1"
  misc_feature
    1. 1641
    /note="wgs contig"
  misc_feature
    15976..157691
    /note="wgs contig"
  misc_feature
    233371..235181
    /note="wgs contig"
  misc_feature
    235232..236508
    /note="wgs contig"

```

ORIGIN

```

Query Match      1.2%; Score 22; DB 2; Length 236508;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 161 AACGAGAAACACTTCGAAA 182
Db 162467 AACGAGAAACACTTCGAAA 162446

```

RESULT 13

```

AC140364 280016 bp DNA linear HTG 05-NOV-2003
LOCUS Mus musculus chromosome UNK clone RP24-75K5, WORKING DRAFT
DEFINITION
SEQUENCE, 8 unordered pieces.
ACCESSION
AC140364.2 GI:38176012
VERSION
HTG; HTGS PHASE1; HTGS_DRAFT; HTGS_FULFILL.
KEYWORDS
Mus musculus (house mouse)
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

REFERENCE

```

1 (bases 1 to 280016)
Wilson, R.K.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 280016)
McPherson, J.D. and Waterston, R.H.
Direct Submission
Submitted (23-FEB-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 280016)
Wilson, R.K.
Direct Submission
Submitted (05-NOV-2003) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Nov 5, 2003 this sequence version replaced gi:28475622.

```

COMMENT

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu

----- Project Information -----

```

Center project name: M_BB0075K05

----- Summary Statistics -----

Sequencing vector: M13; 0%
Chemistry: dye-primer ET; 0% of reads
Chemistry: dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 278028 bases at least Q40
Consensus quality: 278887 bases at least Q30
Consensus quality: 279492 bases at least Q20

* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1202: contig of 1202 bp in length
* 1203 1302: gap of unknown length
* 1303 3866: contig of 2564 bp in length
* 3867 3967: gap of unknown length
* 3967 24768: contig of 20802 bp in length
* 24769 24869: gap of unknown length
* 24869 48433: contig of 23565 bp in length
* 48434 48534: gap of unknown length
* 48534 85803: contig of 37270 bp in length
* 85804 85903: gap of unknown length
* 85904 127685: contig of 41782 bp in length
* 127686 127785: gap of unknown length
* 127786 198320: contig of 70435 bp in length
* 198321 280016: gap of unknown length
* 198321 280016: contig of 81696 bp in length.

FEATURES

SOURCE

1. 280016
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="X"
/clone="RP24-75K5"

misc_feature
/note="assembly_name:Contig20"
1303..3866
misc_feature
/note="assembly_name:Contig22"
3967..24768
misc_feature
/note="assembly_name:Contig23"
24869..48433
misc_feature
/note="assembly_name:Contig24"
48534..85803
misc_feature
/note="assembly_name:Contig25"
85904..127685
misc_feature
/note="assembly_name:Contig26"
127786..198320
misc_feature
/note="assembly_name:Contig27"
198321..280016
misc_feature
/note="assembly_name:Contig28"

ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 280016;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 567 TGTGAGTCAAAAGCCCATGA 588
|||||
Db 58633 TGTGAGTCAAAAGCCCATGA 58654

RESULT 14
AC073813 304407 bp DNA linear HTG 29-JUN-2000
LOCUS AC073813
DEFINITION Mus musculus clone RP23-69B1, WORKING DRAFT SEQUENCE, 58 unordered

ACCESSION
AC073813
VERSION
AC073813.1 GI:8810430
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 304407)
DOE Joint Genome Institute.
TITLE
Sequencing of Mouse
JOURNAL
Unpublished
2 (bases 1 to 304407)
REFERENCE
DOE Joint Genome Institute.
AUTHORS
Direct Submission
JOURNAL
Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 1763393
Center clone name: RP23-69B1

----- Summary Statistics -----

Consensus quality: 257140 bases at least Q40
Consensus quality: 282847 bases at least Q30
Consensus quality: 288639 bases at least Q20
Estimated insert size: 272860; agarose-tp estimation
Estimated insert size: 298707; sum-of-contigs estimation
Quality coverage: 9.46 in Q20 bases; agarose-tp estimation
Quality coverage: 8.64 in Q20 bases; sum-of-contigs estimation.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 58 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 1167: contig of 1167 bp in length
* 1168 1267: gap of unknown length
* 1268 2281: contig of 1014 bp in length
* 2282 2382: gap of unknown length
* 2382 3767: contig of 1386 bp in length
* 3768 3867: gap of unknown length
* 3868 5372: contig of 1505 bp in length
* 5373 5472: gap of unknown length
* 5473 6494: contig of 1022 bp in length
* 6495 6594: gap of unknown length
* 6595 7909: contig of 1315 bp in length
* 7910 8009: gap of unknown length
* 8010 9416: contig of 1407 bp in length
* 9417 9517: gap of unknown length
* 9517 10776: contig of 1260 bp in length
* 10777 10876: gap of unknown length
* 10877 12279: contig of 1403 bp in length
* 12280 12379: gap of unknown length
* 12380 13920: contig of 1541 bp in length
* 13921 14020: gap of unknown length
* 14021 15417: contig of 1397 bp in length
* 15418 15517: gap of unknown length
* 15518 16897: contig of 1380 bp in length
* 16898 16997: gap of unknown length
* 16998 18940: contig of 1943 bp in length
* 18941 19040: gap of unknown length
* 19041 20890: contig of 1850 bp in length
* 20891 20990: gap of unknown length
* 20991 22965: contig of 1975 bp in length
* 22966 23065: gap of unknown length
* 23066 24501: contig of 1436 bp in length

24502 24601: gap of unknown length
 24602 24628: contig of 1827 bp in length
 24629 26528: gap of unknown length
 26529 28503: contig of 1975 bp in length
 28504 28603: gap of unknown length
 28604 30124: contig of 1521 bp in length
 30125 30224: gap of unknown length
 30225 32392: contig of 2068 bp in length
 32393 32392: gap of unknown length
 32393 34644: contig of 2252 bp in length
 34645 34744: gap of unknown length
 34745 37728: contig of 2984 bp in length
 37729 40230: contig of 2402 bp in length
 37829 40230: gap of unknown length
 40331 42284: contig of 1964 bp in length
 40331 42394: gap of unknown length
 42295 44875: contig of 2481 bp in length
 44295 44875: gap of unknown length
 44876 48629: contig of 3654 bp in length
 44876 48729: gap of unknown length
 48630 51512: contig of 2783 bp in length
 48730 51612: gap of unknown length
 51513 54010: contig of 2398 bp in length
 51613 54010: gap of unknown length
 54011 55956: contig of 1846 bp in length
 54111 55956: gap of unknown length
 55957 57509: contig of 1453 bp in length
 55957 57609: gap of unknown length
 57510 63737: contig of 6128 bp in length
 57610 63737: gap of unknown length
 63738 66881: contig of 3044 bp in length
 63838 66881: gap of unknown length
 66882 70918: contig of 3937 bp in length
 66982 70918: gap of unknown length
 70919 75925: contig of 4907 bp in length
 71019 75925: gap of unknown length
 75926 76025: contig of 2915 bp in length
 76026 78941: gap of unknown length
 78941 84797: contig of 5757 bp in length
 79041 84897: gap of unknown length
 84798 88965: contig of 4068 bp in length
 84898 89065: gap of unknown length
 88966 92420: contig of 3355 bp in length
 89066 92420: gap of unknown length
 92421 97335: contig of 4815 bp in length
 92521 97335: gap of unknown length
 97336 102986: contig of 5451 bp in length
 97436 102986: gap of unknown length
 10287 106824: contig of 3838 bp in length
 10287 106824: gap of unknown length
 106825 106924: contig of 7275 bp in length
 106925 114299: gap of unknown length
 114200 114299: contig of 6691 bp in length
 114300 121090: gap of unknown length
 120991 129531: contig of 8441 bp in length
 121091 129531: gap of unknown length
 129532 134893: contig of 5262 bp in length
 129532 134893: gap of unknown length
 134894 142772: contig of 7779 bp in length
 134994 142872: gap of unknown length
 142773 157121: contig of 14249 bp in length
 142873 157121: gap of unknown length
 157122 157221: contig of 7480 bp in length
 157222 164701: gap of unknown length
 164702 171481: contig of 6680 bp in length
 164802 171481: gap of unknown length
 171482 171581: contig of 12427 bp in length
 171582 184008: gap of unknown length
 184009 184108: gap of unknown length
 184109 192392: contig of 8284 bp in length
 192393 203020: gap of unknown length
 192493 203120: gap of unknown length
 203021 203120: gap of unknown length

203121 211871: contig of 8751 bp in length
 211872 211971: gap of unknown length
 211972 223920: contig of 11949 bp in length
 223921 224020: gap of unknown length
 224021 237820: contig of 13800 bp in length
 237821 237920: gap of unknown length
 237921 256627: contig of 18707 bp in length
 256628 256727: gap of unknown length
 256728 280655: contig of 23928 bp in length
 280656 304407: gap of unknown length
 280756 304407: contig of 23652 bp in length.

FEATURES
 SOURCE
 1. 304407
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone_lib="RP23-69B1"
 /clone_lib="RP23-69B1" mouse BAC library 23"

ORIGIN
 Query Match 1.2%; Score 22; DB 2; Length 304407;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 436 GCTCAGCACACGCTGTGCAGG 457
 Db 5763 GCTCAGCACACGCTGTGCAGG 5784

RESULT 15
 AE005048
 LOCUS
 Halobacterium sp. NRC-1 section 79 of 170 of the complete genome.
 DERIVATION
 AE005048 AE004437
 ACCESSION
 AE005048.1 GI:10580743
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Halobacterium sp. NRC-1
 Halobacterium sp. NRC-1
 Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
 Halobacteriaceae; Halobacterium.

REFERENCE
 AUTHORS
 1 (bases 1 to 10225)
 Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
 Shukla,H.D., Laskey,S.R., Balliga,N., Thorsson,V., Shrogha,J.,
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welter,R., Goo,Y.A.,
 Leitbauer,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
 Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angvine,C.M., Dale,H.,
 Isenbarger,T.A., Beck,R.F., Pohlischrodt,M., Spudich,J.L.,
 Uung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
 DasSarma,S.
 Genome sequence of Halobacterium species NRC-1
 Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
 MEDLINE
 PUBMED
 20504483
 11016950
 2 (bases 1 to 10225)
 Ng,W.V., Kennedy,S.P., Mahairas,G.G., Bergquist,B., Pan,M.,
 Shukla,H.D., Laskey,S.R., Balliga,N., Thorsson,V., Shrogha,J.,
 Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Welter,R., Goo,Y.A.,
 Leitbauer,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
 Maddocks,D.G., Jablonski,P.E., Krebs,M.P., Angvine,C.M., Dale,H.,
 Isenbarger,T.A., Beck,R.F., Pohlischrodt,M., Spudich,J.L.,
 Uung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
 Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
 DasSarma,S.
 Direct Submission
 Submitted (14-JUL-2000) Institute for Systems Biology, 4225
 Roosevelt Way NE, Seattle, WA 98105, USA

TITLE
 JOURNAL
 FEATURES
 SOURCE
 1. 10225
 /organism="Halobacterium sp. NRC-1"
 /mol_type="genomic DNA"
 /strain="NRC-1"
 /db_xref="taxon:64091"

gene 125. .1510
/gene="huth"
/note="synonym: VNG1212G"
CDS 125. .1510
/gene="huth"
/note="huth"
/codon_start=1
/transl_table=11
/product="histidine ammonia-lyase"
/protein_id="AAG19580.1"
/db_xref="GI:10580744"
/translation="MSDTRIDADREALQANIYRSHAGAGSELDTAAYRALLVTRLN
ALAKYSGIRERVLVGLNVEVHPVPPVSGSLGASGDLAPLAMSRVLLGEGQAD
VAGERMPAAELAAADLEPVTLOAKEGALINGTQLTGVALALVDAERVRSADTA
GALTFTEVMSSTASCAPAIHEVRPHDGOVSARH1RNLVAGESEVLDHHRDCRVODAY
SRCLPOYHGAVRDALDHLRAVATEINSGATNDPLVPAGDRVGPRASTGDRAAVYSGG
NFRGVLAIRLGYASALAEALAIISERRDRILNPETQEPHLEPFLAPDSGHSGLMT
FOYTNASLVNDLRSLGQPTLDNASVSGQEDHVSAGAIYFRAEVAEKATVAVGVEL
LCGAGREFLEPDLALGAGTAAAYDLVREVSPPVAGDRALADDMAVGDVRAAGVED
AVALADAPLA"
gene 1598. .2578
/gene="VNG1213C"
CDS 1598. .2578
/gene="VNG1213C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1213C"
/protein_id="AAG19581.1"
/db_xref="GI:10580745"
/translation="MSAAQOTVYHRSACRSGRASDIGNPLPTATFDRLHCHATAATR
PTSECAPTERASAPRREIPLMPCOPQAVMRVENSFLPAAGVGERTERLMQHV
THMDFDGGIGIGDATENIYSFIDAERALDAGNTSPFADLPSELMLYENFAD
DVAEPDITGGLDGHSSVYTVYSVHAGGDTTLVNGSDLTARLQALLDAPVAVSGG
KRFDAPLEHHPDLSLDSPHLDLMPCRDLDTGGLKOVERDVGVRASVDDVDRGA
VRLMRYABDNDEALDRLVYNYQDTQNLQALIDAVARLHRDVEEPHV"
gene 3236. .4366
/gene="VNG1214H"
CDS 3236. .4366
/gene="VNG1214H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1214H"
/protein_id="AAG19582.1"
/db_xref="GI:10580746"
/translation="WTSHTDRLAEIEDGPELRALIDLTTRDRIQGVDPQGRKDG
KDGGLDGPDSNTIIVHSRREDKQKLTDLGAADHDROYDI FVYVNTNIIITGNQ
KVPDVAQPFVDEYGEIDIMDGERLSELNHHQDLREYLR1ARDEPTEKTAARLI
DERLRLRRDNELEPRVADAPTAVLHLVPHDAVSGDUDFRHDLPTSLPGRYGGYS
YENTLDGRVTVYARGGDNEPDREGIYIVDTGEGWEADAMNGENMTTIGGOSPEKLI
AYEYGRNALNELELDGPFVEGLSVISYKGISATKKGCFDNGPKVRONDIEAKPH
TVEBDSPTEGMRKGFDRVMKARWSDGSPYSIDNEWHEFER"
gene 4366. .4638
/gene="trn19"
tRNA complement(4565. .4638)
/gene="trn19"
/product="tRNA-Thr"
complement(4650. .5132)
/gene="pai1"
/note="synonym: VNG1215G"
complement(4650. .5132)
/gene="pai1"
/note="pai1"
/codon_start=1
/transl_table=11
/product="sporulation regulator homolog"
/protein_id="AAG19583.1"
/db_xref="GI:10580747"
/translation="WTVRPAAGSTDDAVALMVLQAAAGORAGSHLLAENNAQARDL
IGQYHADVAVTTQAGRTVGVDMHTGTGFEYDTRTGAVINDLTVAGPFGDGGSA
LITHAERELRSRGAALAI EALADNAAARLRYERGVA PHRTFEOSEVNDTHSKDP
"

gene complement(5129. .6325)
/gene="pgk"
/note="synonym: VNG1216G"
CDS complement(5129. .6325)
/gene="pgk"
/note="pgk"
/codon_start=1
/transl_table=11
/product="glucose-6-phosphate isomerase"
/protein_id="AAG19584.1"
/db_xref="GI:10580748"
/translation="WAIRTLDDLAANRAIGRVINDINSEPLTAGGLADBARLRAVDT
LAELLAADARAVVALHQGRPGDEPARLERHADRLDALDPAVSYCDATFFSGARAV
ADLAGEVIVLENTFYESEYWAFAPERAADTALVDGLADALAYNDAFAAHRSQP
SLVGPETLPSYAGWEAEALDLSGVAADTPERTPTVYVGAAPVSPSEVAARALSHGL
ADNVLVGVNINFLAAGVNDGRASSTDPIHERDYGTEIARAADILAAHNDALHL.PVD
VAVERDGRCEIETALPPAGDEAVCDIGSDTVDAAYAVYLADETYVYVNGPAGVEDD
LFADTRKVPFPAASVEHSHVGGGDTAAIRRFDTITGFDHSTGGGAALNLTLDALP
AVAAALR"
gene complement(6428. .7144)
/gene="VNG1218C"
CDS complement(6428. .7144)
/gene="VNG1218C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1218C"
/protein_id="AAG19585.1"
/db_xref="GI:10580749"
/translation="WQRFAGCTPTA VMAVALGPPRPSVHGFYRGLYRKMNDVLEB
ALTSARVVSASAPDPAATLREHDATAAVVORAGDLAGVTADAVDAVADGGLD
SVTVADVNAADPVSLSVATVADADIMARTRDGCVLVVDADABAIIVASADALALA
ATNDORSRBPILGGSDRTPVAGASGAGDPPEHSDGPTTAGANTYSBGQGCERCGG
LAALGRNGRLCTECSL"
gene 7107. .7853
/gene="urk"
/note="synonym: VNG1219G"
CDS 7107. .7853
/gene="urk"
/note="urk"
/codon_start=1
/transl_table=11
/product="uridine kinase"
/protein_id="AAG19586.1"
/db_xref="GI:10580750"
/translation="WTAAGVOPARCIQHASPPRAVSMTIS9FAIGAGTGAGKTV
AREITDNGESATLPLDNYTEDSDRPEERANANYDHPSEFELRTHIDALLSG
QSIEHPQYDFERHAKADRVVPEPTDVLEGIALLSDETVMDMLDHLIYVETDADR
ILRIERDVVERGELGEMVOQYSTVPMHQFLEPTKRHADIIIPGANSVAVNLL
BEKVOAESTEMTAMARDDTAFEDATVNPQ"
gene complement(7911. .8234)
/gene="VNG1220H"
CDS complement(7911. .8234)
/gene="VNG1220H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1220H"
/protein_id="AAG19587.1"
/db_xref="GI:10580751"
/translation="NAEMEDPSGRDWVHDPGEHSGEGGRKYGLAVLSKMTVEGDFP
LDAAEFVAEYGEYPRLNHQTVSVADIFHHIDADEPTKVGPHKTVGAALQDDADFLD
YHPDA"
gene complement(8533. .8615)
/gene="trn20"
tRNA complement(8533. .8615)
/gene="trn20"
/product="tRNA-Ser"
8926. .10119
/gene="orc8"
/note="synonym: VNG1224G"
8926. .10119
/gene="orc8"

```
/note="Orc8"  
/codon_start=1  
/transl_table=1  
/product="Orc / cell division control protein 6"  
/protein_id="AA019588.1"  
/db_xref="GI:10580752"  
/translation="MKTFRRRVRLFANKDVLKDHVPEEREREREIDQYANALQDV  
VDGMEPDNVFVYGTGKTVATRIMDALFEYDADDDGVDSTSVENCHHPSXQ  
AAIALVNEIRGDTSDPLTGLSTSDVINALFDEIARBGTVLILDEIDNLDDDDL  
LYOLPRAKTNIGIEDSQVAVVGIENDYTFRNDLSPKVQDTLCEREIKFPYPYDANEIYT  
IIDDRAERALSGLVLTGAVIPQCAALAAARDGSARQALDILRESVNAIEDERTIVE  
DVEYAVRRVERGRIKDSIKDLTHGQVVLAVTQTAIADDTPVRAKELYEVYADIAA
```

Query Match	1.1%;	Score 21;	DB 1;	Length 10225;
Best Local Similarity	100.0%;	Pred. No. 7.9;		
Matches	21;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY	1104	GGCCGAACGCTCCACCGGGA	1124	
Db	2536	GGCCGAACGCTCCACCGGGA	2556	

Search completed: March 30, 2004, 14:48:31
Job time : 7035.56 secs

TITLE		JOURNAL		COMMENT	
Direct Submission		Submitted (01-JUL-2003) Weissshaar B., Max-Planck-Institut fuer Zuechtungsforchung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany		This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone M0M1. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/ .	
FEATURES		source		Location/Qualifiers	
		1..395		/organism="Arabidopsis thaliana"	
				/mol type="genomic DNA"	
				/strain="Columbia 0"	
				/db xref="taxon:3702"	
				/clone="GK-547E01-020631"	
				/clone_11b="Arabidopsis thaliana T-DNA insertion lines"	
				/note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"	
ORIGIN		Query Match		52.5%; Score 364.6; DB 29; Length 395;	
		Best Local Similarity		95.2%; Pred. No. 2.1e-50;	
		Matches		376; Conservative 0; Mismatches 19; Indels 0; Gaps 0;	
QY	114	ACTGATGAAATTGGAGTCAGAAATCGAAAAACGAGCGCTTTAGACTTAATAG	173		
Db	1	ACTGATGAAATTGGAGTCAGAAATCGAAAAACGAGCGCTTTAAAGCTTAATAG	60		
QY	174	CTTCTCATTTGTCTCTTCTTGTCGATTTATTTCTCCCGGAGTCGATCACTA	233		
QY	61	CTTCTCATTTGTCTCTTCTTGTCGATTTATTTCTCCCGGAGTCGATCACTA	120		
Db	234	CTTCTCATTTGTCTCTTCTTGTCGATTTATTTCTCCCGGAGTCGATCACTA	293		
QY	121	CTTCTCATTTGTCTCTTCTTGTCGATTTATTTCTCCCGGAGTCGATCACTA	180		
QY	294	CTTGAAGCCTCCGATGCGCTCAACGATGATTTCTGCTGATTTCTTTTCTGCG	353		
Db	181	CTTGAAGCCTCCGATGCGCTCAACGATGATTTCTGCTGATTTCTTTTCTGCG	240		
QY	354	TGAAAAAATTTGCCCTAATGTTTCGATTTGAAAGTTTGGCTATGGTTACTTTT	413		
Db	241	TGAAAAAATTTGCCCTAATGTTTCGATTTGAAAGTTTGGCTATGGTTACTTTT	300		
QY	414	CCCTATATTTTAAAGTCTTAAAGTAAAGATCCGATGATTTGTTGTTGCTATTTG	473		
Db	301	CCCTATATTTTAAAGTCTTAAAGTAAAGATCCGATGATTTGTTGTTGCTATTTG	360		
QY	474	TTGTGCTTTTCAACGTTTAAAGTAAAGATCCGATGATTTGTTGTTGCTATTTG	508		
Db	361	TTGTGCTTTTCCCATTTTGTGCTGATGACAGAAAT	395		
RESULT 2		BX547724		303 bp DNA linear GSS 02-JUL-2003	
LOCUS		Arabidopsis thaliana T-DNA flanking sequence GK-547E01-020586,			
DEFINITION		genomic survey sequence.			
ACCESSION		BX547724			
VERSION		BX547724.1			
KEYWORDS		GSS.			
SOURCE		Arabidopsis thaliana (thale cress)			
		Arabidopsis thaliana			
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			

REFERENCE AUTHORS TITLE	JOURNAL
1 Strizhov,N., Li,Y., Rosso,M., Viehoveer,P., Dekker,K., Saedler,H. and Weishaar,B. A pipeline for automated high-throughput generation of FSTs (flanking sequence tags) from Arabidopsis thaliana T-DNA transformed lines Unpublished	2
Rosso,M., Strizhov,N., Li,Y., Reiss,B., Dekker,K. and Weishaar,B. A new Arabidopsis thaliana T-DNA mutagenised population (GABI-Kat) for flanking sequence tag based reverse genetics Unpublished	3 (bases 1 to 303)
Rosso,M., Li,Y., Strizhov,N. and Weishaar,B. Direct Submission Submitted (01-JUL-2003) Weishaar B., Max-Planck-Institut fuer Zuechtungsforshung, Carl-von-Linne-Weg 10, Koeln, 50829, Germany This sequence is recovered from the left border of the T-DNA. It indicates an insertion within the locus defined by clone W0M1. The sequences are generated at the MPI for Plant Breeding Research in the context of the GABI-Kat project. GABI-Kat is part of the German Plant Genomics program designated 'GABI'. Information on line availability can be found at: http://www.mpiz-koeln.mpg.de/GABI-Kat/ location/Qualifiers	1..303
FEATURES source	<p>1..303</p> <p>/organism="Arabidopsis thaliana" /mol_type="Genomic DNA" /strain="Columbia 0" /db_xref="taxon:3702" /clone="GK-547E01-020586" /clone_11b="Arabidopsis thaliana T-DNA insertion lines" /note="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were removed"</p>
ORIGIN	
Query Match 39.9%; Score 277.4; DB 29; Length 303; Best Local Similarity 94.7%; Pred. No. 56-36; 16; Indels 0; Gaps 0; Matches 287; Conservative 0; Mismatches 16; Indels 0; Gaps 0;	
QY 115 CTGAATGAAATTTGGAGTCAGAAATCGAAAAAGAGCCGCTTTTAAAGCTTAATAGC 174	
Db 1 CTGAATGAAATTTGGAGTCAGAAATCGAAAAAGAGCCGCTTTTAAAGCTTAATAGC 60	
QY 175 TTTCCTATTGCTCTCTTCGTCGAGTTATTTCTTCCTCCGAGTCTGACTCACTAC 234	
Db 61 TTTCCTATTGCTCTCTTCGTCGAGTTATTTCTTCCTCCGAGTCTGACTCACTAC 120	
QY 235 TCTCACTCTCCGCGCTTTAACTTAAGTTCCTCGCTGTTTACTGTAAAGTTTCTGCC 294	
Db 121 TCTCACTCTCCGCGCTTTAACTTAAGTTCCTCGCTGTTTACTGTAAAGTTTCTGCC 180	
QY 295 TTAAAGCTTCGATCGCTTCACCGCATCGATTCGTGTCGATTCCTCTTTTCTTCGCT 354	
Db 181 TTAAAGCTTCGATCGCTTCACCGCATCGATTCGTGTCGATTCCTCTTTTCTTCGCT 240	
QY 355 GGAATAATGCTTATGTTCTCGATTCGAAAGTTTGTGTATGAGTTACTTTTTC 414	
Db 241 GGAATAATGCTTATGTTCTCGATTCGAAAGTTTGTGTATGAGTTACTTTTTC 300	
QY 415 CCT 417	
Db 301 CCT 303	

RESULT 3
BH854770
LOCUS
DEFINITION BH854770 209 bp DNA linear GSS 08-JUL-2002
SAUK_088648.19.55.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SAUK_088648.19.55.x, genomic
survey sequence.
ACCESSION
BH854770
VERSION
BH854770.1 GI:21704360
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 209)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadriab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shim,P., Zimmerman,J. and Ecker,J.R.
A Sequence-indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
JOURNAL
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (Signal)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckersalk.edu
This is single pass sequence recovered from the left border of
TDNA.
FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..209
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SAUK_088648.19.55.x"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN
Query Match 22.7%; Score 157.8; DB 28; Length 209;
Best Local Similarity 84.7%; Pred. No. 2.6e-16;
Matches 177; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 322 GCATTCTGCTCGATTCTCTTTTCTGCGTGAATAATGCCCTAATGTTTCGATT 381
DB 1 GCATCTGTCGCGATTTCGCTTTTGTTCACAGAAATATGCCCTAATGTTTCGATT 60
QY 382 TCGAAGCTTTTTCGCTAATGCTTCTTTTCCCTAATTTAATAGTTCTTAAGTAACG 441
DB 61 TCGAAGCTTTTTCGCTAATGCTTCTTTTCCCTAATTTAATAGTTCTTAAGTAACG 120
QY 442 ATACCTGCGCTCTTACTGTTTGTTCATTGTTGCTTTCACCGTTTAACTGCGTGAATC 501
DB 121 ATGCTGCGCTCTTACTGCTAATGACAGTTCGCGGTGCTTCCCGTTAAGCCGCTGATC 180
QY 502 GAGATATTGACTGTAATAATCCTTCGT 530
DB 181 GAGATATTGACTGTAATAATCCTTCGT 209
RESULT 4
BH847287
LOCUS
DEFINITION BH847287 144 bp DNA linear GSS 13-JUN-2002
SAUK_050870.18.25.x Arabidopsis thaliana TDNA insertion lines

ACCESSION
BH847287
VERSION
BH847287.1 GI:21418158
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 144)
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R.,
Gadriab,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L.,
Shim,P., Zimmerman,J. and Ecker,J.R.
A Sequence-indexed Library of Insertion Mutations in the
Arabidopsis Genome
Unpublished (2001)
JOURNAL
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (Signal)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckersalk.edu
This is single pass sequence recovered from the left border of
TDNA.
FEATURES
source
Class: TDNA tagged.
Location/Qualifiers
1..144
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SAUK_050870.18.25.x"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna_protocols.html"
ORIGIN
Query Match 15.7%; Score 108.8; DB 28; Length 144;
Best Local Similarity 84.7%; Pred. No. 3.6e-08;
Matches 122; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
QY 372 GTTCTCGATTTCGAGGTTTGTGCTAATGGCTTACTTTTCCCTAATTTTAATAGTTC 431
DB 1 GTTCTCGAGTTTCGAGGATTGTTGCTAGGCGCTACTTTTCCCTAATTTTAACAGTC 60
QY 432 TTAGTAACGATCTGCGCTTACTGTTTGTTCATTGTTGCTCTTACCGCTTAA 491
DB 61 TTAGTAACGATCTGCGCTTACTGTTTGTTCATTGTTGCTCTTACCGCTTAA 120
QY 492 GTTCGTAATCGAGATTTGACTG 515
DB 121 GTTCGTAATCGAGATTTGACTG 144
RESULT 5
AQ964581
LOCUS
DEFINITION AQ964581 232 bp DNA linear GSS 28-JAN-2000
LERGX20TR LERG Arabidopsis thaliana genomic clone LERGX20, genomic
survey sequence.
ACCESSION
AQ964581
VERSION
AQ964581.1 GI:6792282
KEYWORDS
GSS.
SOURCE
Arabidopsis thaliana (thale cress)
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (bases 1 to 232)
 AUTHORS Busell, C.R., Lin, X., Pat, G., Barnstead, M., Bowman, C., Uterbach, T.,
 TITLE Feildblum, T., Liang, F., Creasy, T., and Fraser, C.M.
 Genomic survey sequencing of Landsberg erecta ecotype of
 Arabidopsis thaliana and identification of sequence-based
 polymorphisms
 JOURNAL Unpublished (2000)
 COMMENT Contact: Xiaoying Lin
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: atc@tigr.org
 For additional information, see <http://www.tigr.org/tdb/at.html>
 Seq primer: TR
 Class: shotgun.

FEATURES
 source Location/Qualifiers
 1..232
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="LANDSBERG ERRECTA"
 /db_xref="taxon:3702"
 /clone_11b="LERGX20"
 /clone_11b="LERG"
 /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
 sheared to 0.4-0.7 kbp before ligation."

ORIGIN
 Query Match 14.9%; Score 103.4; DB 28; Length 232;
 Best Local Similarity 94.7%; Pred. No. 2e-07; Indels 0; Gaps 0;
 Matches 107; Conservative 0; Mismatches 6;

Qy 583 TGTGTTGTTTGGAGCCGATGCGTGTGGCTGTATTAACCTTCAGCTTCATGTGAT 642
 Db 16 TCGCGAAGTTGAGCCATGCGTGTGGCTGTATTAACCTTCAGCTTCATGTGAT 75
 Qy 643 TTTGAGTTTGTGATGACGTGCGTTCTTTGGTGGCTATAGTTGTA 695
 Db 76 TTTGAGTTTGTGATGACGTGCGTTCTTTGGTGGCTATAGTTGTA 128

RESULT 6
 CF872119 706 bp mRNA linear EST 31-OCT-2003
 LOCUS tric029xp16.b1 T. reesei mycelial culture, Version 6 October 2003
 DEFINITION Hypocrea jecorina cDNA clone tric029xp16, mRNA sequence.
 ACCESSION CF872119
 VERSION CF872119.1 GI:38126801
 KEYWORDS EST.
 SOURCE Hypocrea jecorina (anamorph: Trichoderma reesei)
 ORGANISM Hypocrea jecorina
 Rukavotsa; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreaceae; Hypocreales; Hypocreaceae; Hypocrea.
 1 (bases 1 to 706)
 Diener, S.E., Dunkel-Schetter, L., Dunn-Coleman, N., Houfek, T.D.,
 Mitchell, T.K., van Solingen, P., Teunissen, P.J.M., Ward, M., and
 Dean, R.A.
 Analysis of the protein processing and secretion pathways in a
 Trichoderma reesei EST dataset
 JOURNAL Unpublished (2003)
 COMMENT Contact: Ralph A. Dean
 Fungal Genomics Laboratory
 North Carolina State University
 Campus Box 7251, Raleigh, NC 27695, USA
 Tel: 919-513-0020
 Fax: 919-513-0024
 Email: ralph.dean@ncsu.edu
 Seq primer: Tr-F1 primer.
 Location/Qualifiers
 1..706
 /organism="Hypocrea jecorina"
 /mol_type="mRNA"
 /strain="QM6a"

ORIGIN
 Query Match 13.2%; Score 92; DB 14; Length 706;
 Best Local Similarity 45.9%; Pred. No. 6.9e-06; Indels 0; Gaps 0;
 Matches 308; Conservative 0; Mismatches 363;

Qy 5 AACAAACAAAATTAGAGATGATGTCGATGAGATTAATTAATAGTGGACATTA 64
 Db 4 AA 63
 Qy 65 GTTACGCAAAAGAGAAAAAGATACAAAATGAAACAAAATCAACTGAATGA 124
 Db 64 AA 123
 Qy 125 ATTGAGATCGAATCGAAGAAAGAGGCGTTTAAAGCTTAATTAAGCTTCAT 184
 Db 124 AAAAGGGGGGGGGGGGTTTCATTTTCTTTTCTTTTCTTTTCTTTTCTTTT 183
 Qy 185 GTCCTCTCTGTCAGTTATTTCTTCTCGAGATCTGACACACACACACAC 244
 Db 184 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 243
 Qy 245 CGGCGCTTAACTTACGTTCCGCTGCTTCTGATGTTCTGCTTACGCTG 304
 Db 244 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 303
 Qy 305 CGATGCGCTACGATGATCTGTCGATTTCTTTTCTTCTGCGTGAATAATG 364
 Db 304 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 363
 Qy 365 CCTTAATGTTCTGATTTGCAAGTTTGGGATGAGGTTTACCTTATTTTCC 424
 Db 364 CTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 423
 Qy 425 ATAGTTCTAGTACGATACCTGCTTACGTTTGTGATTTGTGTCATTTGTC 484
 Db 424 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 483
 Qy 485 CCGTTAGTGTGATCGAGATTTGACGTGAAATCGTGTGTTTGGTTTGT 544
 Db 484 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 543
 Qy 545 TTCAATTAATGAGATGATACCTTTGCTTGAATGTTGTTTGAAGCTATG 604
 Db 544 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 603
 Qy 605 GTTGTGCTGTTTAACTTACGTTCAAGTTGATGTTGAGTTTGGATGAC 664
 Db 604 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 663
 Qy 665 TGGGTTTCTT 675
 Db 664 CTTTCTTTTCTT 674

RESULT 7
 CNS0073W 922 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence TERS end of BAC #
 DEFINITION BAC14D09 of RFLP-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL066784
 VERSION AL066784.1 GI:4945247
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 922)
GENOSCOPE.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
location/Qualifiers
1..922
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/note="end : TET3"
FEATURES
source
Query Match 12.7%; Score 88.4; DB 29; Length 922;
Best Local Similarity 29.4%; Pred. No. 2.3e-05;
Matches 200; Conservative 165; Mismatches 309; Indels 6; Gaps 2;
ORIGIN
2 AAAAAAATAAATTAAGCAAGTATGTCGATTAATTAATTAAGTGAACAT 61
|||||
239 AAAAAAATAAATTAAGCAAGTATGTCGATTAATTAATTAAGTGAACAT 298
|||||
62 TAAGTTAAGCAAGTATGTCGATTAATTAATTAAGTGAACAT 121
|||||
299 CAAAAAATAAATTAAGCAAGTATGTCGATTAATTAAGTGAACAT 358
|||||
122 AAAAAAATAAATTAAGCAAGTATGTCGATTAATTAAGTGAACAT 181
|||||
359 AAAAAAATAAATTAAGCAAGTATGTCGATTAATTAAGTGAACAT 413
|||||
182 TTGTGCTCTCTGTCGATTAATTAAGTGAACAT 241
|||||
414 TTGTGCTCTCTGTCGATTAATTAAGTGAACAT 473
|||||
242 TTGTGCTCTCTGTCGATTAATTAAGTGAACAT 300
|||||
474 TTGTGCTCTCTGTCGATTAATTAAGTGAACAT 533
|||||
301 CCTCGATGCTCAACCGCATTTCTGCTGATTTCTTTTCTGCTGAAAA 360
|||||
534 GGTGKGGTCTTTTSTGTGKGGTCTTTTSTGTGKGGTCTTTTSTGTG 593
|||||
361 ATTGCGCTATGTTCTGATTTGCAAGTTTGTGCTATGCTTTTCTTANA 420
|||||
594 TGGKGGKGGTCTTTTGTGKGGTCTTTTGTGKGGTCTTTTGTGKGG 653
|||||
421 TTGTAGTCTTGAAGTATGATTTGATTTGATTTGATTTGATTTGAT 480
|||||
654 TTGGGKGGTCTTTTGTGKGGTCTTTTGTGKGGTCTTTTGTGKGG 713
|||||
481 TTACCGTTAGTCTGATGAGATTTGATTTGATTTGATTTGATTTG 540
|||||

Db 714 TTK 773
Oy 541 TTGTTTATATATATGATGATGATGATGATGATGATGATGATGAT 600
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 774 TTK 833
Oy 601 ATGCGTTGTCGTTGTTTAACTTCACTGATGATGATGATGATGAT 660
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 834 GTTKGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 893
Oy 661 ACTGCGGTTTCTTTGTCG 680
:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 894 KKK 913
RESULT 8
CNS005TE
LOCUS 997 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL060767.1 GI:4943573
VERSION 1
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 997)
GENOSCOPE.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazuo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
location/Qualifiers
1..997
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/note="end : TET3"
FEATURES
source
Query Match 12.7%; Score 88; DB 29; Length 997;
Best Local Similarity 20.8%; Pred. No. 2.5e-05;
Matches 115; Conservative 160; Mismatches 278; Indels 0; Gaps 0;
ORIGIN
42 AAATTAATGTCGAAATTAATTAATTAATTAATTAATTAATTAATTA 101
|||||
335 ATNTAANNANNTAATTAATTAATTAATTAATTAATTAATTAATTA 394
|||||
102 AAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 161
|||||
395 AAAAAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 454
|||||
162 GAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 221
|||||

Matches	54; Conservative	185; Mismatches	129; Indels	0; Gaps	0;
QY	324	ATTCTGTCGTGATTTCTCTTCTTGCGTGGAAAAATGCCCTAATGTTCTGCATTC	383		
Dd	330	MMMTMMTT	389		
QY	364	CAGAGTTTTGTGCGTAAGGGTACTTTTTCCCTAATTTATAAGTCTTAGTAACGAT	443		
Dd	390	TTTTTTTTTTTTTKTKTKTTTTTTTTTTTTTTTTTTTTTTTTTKTKTKTKTKTKKKKKT	449		
QY	444	ACGTCGCTTACGTTTGTTCATTTGTGTGTCCTTCAACCCTTAGTGCGCGANCG	503		
Dd	450	KTTTGTKTKKKKKTKTKTKTKTKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK	509		
QY	504	AGTATTTCAGTCGTAAAAAACCTTCGTTTTTNGTTTTGTTTCATATAAATCGATTGA	563		
Dd	510	KKK	569		
QY	564	TCTACCTTTTNGCCTTGTGATGTGTTTTTTTGAAGCTAAGCGTTGTGCTGTGTAAC	623		
Dd	570	KK	629		
QY	624	TTTCAGCTTCAGTCGTGATTTTGAAGTTTGTAGTAGCATGTGGAGTTCTTGTCGCTA	683		
Dd	630	KK	689		
QY	684	TAGTTGT	691		
Dd	690	KKKKKKKK	697		

RESULT 10	613 bp	DNA	linear	GSS 28-JAN-2000
AO964580				
LOCUS				
DEFINITION	LERGX20TF LERGX Arabidopsis thaliana genomic clone LERGX20, genomic			
ACCESSION	survey sequence.			
VERSION	AO964580			
KEYWORDS	AO964580.1 GI:6792281			
SOURCE	GSS.			
ORGANISM	Arabidopsis thaliana (thale cress)			
	Arabidopsis thaliana			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
	rosids; eurosids II; Brassicales; Brassicaceae; Arabidopses.			
REFERENCE	1 (baes 1 to 613)			
AUTHORS	Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,			
	Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.			
TITLE	Genomic survey sequencing of Landsberg erecta ecotype of			
	Arabidopsis thaliana and identification of sequence-based			
	polymorphisms			
JOURNAL	Unpublished (2000)			
COMMENT	Contact: Xiaoying Lin			
	The Institute for Genomic Research			
	9712 Medical Center Dr., Rockville, MD 20850, USA			
	Tel: 301 838 0200			
	Fax: 301 838 0208			
	Email: at@tigr.org			
	For additional information, see http://www.tigr.org/tdb/at/at.html			
	Seq primer: TF			
	Class: shotgun.			

ORIGIN	12.28;	Score 84.8;	DB 28;	Length 613;
Query Match				

Best Local Similarity 97.7%; Pred. No. 0.00012;
Matches 86; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 608 GTTGGCTTTAACTTCAACGCTGATGATTTGAGTTTGTAGTACGCTG 667
Db 613 GTTGGATGTTAACTTCAACGCTGATGATTTGAGTTTGTAGTACGCTG 554

QY 668 GTTCTTTGTGCTATAGTTGTAA 695
Db 553 GTTCTTTGTGCTATAGTTGTAA 526

RESULT 11
LOCUS AU236368 647 bp mRNA linear EST 01-APR-2002
DEFINITION AU236368 RAF14 Arabidopsis thaliana cDNA clone RAF14-93-K05 5',
MRNA sequence.
ACCESSION AU236368.1 GI:19875537
VERSION AU236368.1
KEYWORDS EST.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 647)
Seki, M., Narusaka, M., Ishida, J., Kamiya, A., Satou, M., Nakajima, M.,
Oono, Y., Sakurai, T., Carninci, P., Kawai, J., Itoh, M., Ishii, Y.,
Arakawa, T., Shibata, K., Shinagawa, A., Mutamatsu, M., Hayashizaki, Y.,
and Shinozaki, K.
Large scale analysis of Arabidopsis full-length cDNA (2002b)
Unpublished (2002)
Contact: Motoaki Seki
Plant Functional Genomics Research Group
RIKEN Genomic Sciences Center
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-4359
Fax: 81-298-36-9060
Email: msek@rtc.riken.go.jp
An Arabidopsis full-length cDNA library was constructed essentially
as reported previously (Seki et al., 1998). cDNA cleaved with BamHI
and XhoI was ligated to modified lambda FL-C1 vector (Carninci et
al., submitted for publication) digested with BamHI and SalI. This
clone is in a modified pBluescript vector. Please visit our web
site: (http://www.gsc.riken.go.jp/e/plant/index_e.html) for further
details.

FEATURES
source Location/Qualifiers
1..647
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="RAF14-93-K05"
/issue_type="root"
/lab_host="DH10B"
/clone_lib="RAF14"
/note="Site_1: BamHI; Site_2: SalI"

ORIGIN
Query Match 12.1%; Score 84; DB 9; Length 647;
Best Local Similarity 100.0%; Pred. No. 0.00015;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 AATTATTTCTTCCCGGAGTCCGACTCACTCACTCTCCGCGCTTTAACT 258
Db 1 AATTATTTCTTCCCGGAGTCCGACTCACTCACTCTCCGCGCTTTAACT 60

QY 258 TACGTTCTCCGCTTACTCTGT 282
Db 61 TACGTTCTCCGCTTACTCTGT 84

RESULT 12
B2558518

LOCUS B2558518 1872 bp DNA linear GSS 17-DEC-2002
DEFINITION pa98401.209.x1 pac82-164 Pseudomonas aeruginosa genomic clone
pa98401.209, genomic survey sequence.
ACCESSION B2558518
VERSION B2558518.1 GI:27173068
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1872)
Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Haestings, M.,
Burns, J.L., Kaul, R. and Olsen, M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: Shotgun.

FEATURES
source Location/Qualifiers
1..1872
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pa98401.209"
/clone_lib="pac82-164"
/note="clinical isolate 2-164 whole genomic shotgun
library."

ORIGIN
Query Match 11.9%; Score 82.8; DB 28; Length 1872;
Best Local Similarity 46.2%; Pred. No. 0.00012;
Matches 304; Conservative 0; Mismatches 348; Indels 6; Gaps 1;

QY 2 ACAAACAACAAATTAAGCAAGTCATGTCGATTAATTAATGCGAACAAT 61
Db 527 AA 586

QY 62 TAACTTAAGCGAAAGAAAGAAAGTACAAATGAACAAATCAATGAATG 121
Db 587 AA 646

QY 122 AAAAAATTTGAGTCAGAAATCGAAAAACGAGCGCTTTAGAGCTTAATAGCTTCCTCA 181
Db 647 AAAAAAGGGGGGAAAAAAGGGGGGGGTTTGGGGGGTCTTTTCTTTT 706

QY 182 TTGCTCTCTTCTGCTGAGTTATTTCTCTCCGAGTCTGACTCACTCTCACT 241
Db 707 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 766

QY 242 CTCGCGCGCTTAACTTAAGTTCCTCCGCTTACTCTGTAAGTTTCTGCTTAGAC 301
Db 767 TTGTGTGTT 826

QY 302 CTCGATCGCCCTACCGAGATTCGTGTCGATTTCTCTTTCTCTGCTGAAAAA 361
Db 827 TTTTCTTTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGTGTTTTTTTGT 886

QY 362 TTGCGCTAATGTTCTGATTTGAAAGTTTGTGCTATGAGTTATTTTCCCTAAT 421
Db 887 TTNGTTTTTGTGTTTTTTTTTTTTTTTGTGTTTTTGTGTTTTTGTGTT 946

QY 422 TTTATAGTCTTAGTAAGATACCTGCGCTTACTGTTTGTTCATTTGTGTCGT 481
Db 947 TTTTATTTGTTTGTGTTTTTTTTTTTTTTTGTGTTTTTGTGTTTTGT 1006

QY 482 TCACGTTTAGTCGCTGATCGAGATTTGAATGAAATAATCTTGTGTTTTT 541

[illegible]

COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGMA)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: eckersalk.edu
 This is single pass sequence recovered from the left border of
 TDNA.
 Class: TDNA tagged.
 Location/Qualifiers
FEATURES
 source
 1..81
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia 0"

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 16:31:11 ; Search time 52.2914 Seconds
(without alignments)
7375.802 Million cell updates/sec

Title: US-10-030-829-1_COPY_1_695
Perfect score: 695
Sequence: 1 gacaacaacaataa.....gtggctatagtgtaaa 695

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*

1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85.6	12.3	7218	1 US-08-232-463-14	Sequence 14, Appl
2	52.6	7.6	7286	3 US-09-331-581-3	Sequence 3, Appl
3	52.6	7.6	7938	3 US-09-331-581-3	Sequence 14, Appl
4	50.6	7.3	396	4 US-09-640-173-53	Sequence 53, Appl
5	50.6	7.3	396	4 US-09-713-550-53	Sequence 4, Appl
6	50.6	7.3	10619	4 US-10-204-708-4	Sequence 7, Appl
7	50.2	7.2	5562	4 US-10-204-708-63	Sequence 63, Appl
8	49.4	7.1	19124	2 US-08-487-826B-13	Sequence 13, Appl
9	47.8	6.9	2447	2 US-09-014-969-14	Sequence 14, Appl
10	47.4	6.8	5666	4 US-10-204-708-29	Sequence 29, Appl
11	47.4	6.8	8607	4 US-10-204-708-72	Sequence 72, Appl
12	47	6.8	8961	4 US-10-204-708-75	Sequence 75, Appl
13	46.8	6.7	19233	4 US-10-204-708-45	Sequence 45, Appl
14	46	6.6	7218	1 US-08-232-463-14	Sequence 17, Appl
15	45.6	6.6	289	3 US-09-007-005-17	Sequence 17, Appl
16	45.6	6.6	289	3 US-09-244-796-17	Sequence 17, Appl
17	45.4	6.5	11049	4 US-10-204-708-23	Sequence 23, Appl
18	44.2	6.4	1493	1 US-08-340-820-24	Sequence 24, Appl
19	44.2	6.4	1493	1 US-08-593-535-24	Sequence 24, Appl
20	44	6.3	240	1 US-08-628-417-6	Sequence 33, Appl
21	44	6.3	2394	4 US-09-800-729-33	Sequence 33, Appl
22	43.6	6.3	359	4 US-09-621-976-16008	Sequence 16008, A
23	43.6	6.3	2915	4 US-09-336-115C-5	Sequence 36, Appl
24	43.6	6.3	5501	4 US-10-204-708-38	Sequence 38, Appl
25	43.4	6.2	140	1 US-08-628-417-5	Sequence 5, Appl
26	43.4	6.2	359	4 US-09-621-976-16019	Sequence 16019, A
27	43.4	6.2	362	4 US-09-621-976-16010	Sequence 16010, A

28	43.4	6.2	365	4 US-09-621-976-16042	Sequence 16042, A
29	43.4	6.2	5455	4 US-10-204-708-33	Sequence 33, Appl
30	43.4	6.2	11015	4 US-10-204-708-55	Sequence 55, Appl
31	43	6.2	578	4 US-09-602-877A-95	Sequence 95, Appl
32	43	6.2	8607	4 US-10-204-708-72	Sequence 72, Appl
33	43	6.2	8961	4 US-10-204-708-75	Sequence 75, Appl
34	43	6.2	19124	2 US-08-487-826B-13	Sequence 13, Appl
35	42.8	6.2	1114	4 US-09-152-060-41	Sequence 41, Appl
36	42.8	6.2	11050	4 US-10-204-708-85	Sequence 85, Appl
37	42.8	6.2	11331	4 US-10-204-708-27	Sequence 27, Appl
38	42.6	6.1	240	1 US-08-628-417-6	Sequence 6, Appl
39	42.6	6.1	371	4 US-09-621-976-16048	Sequence 16048, A
40	42.6	6.1	10144	4 US-10-204-708-94	Sequence 94, Appl
41	42.4	6.1	270	2 US-08-520-678A-10	Sequence 30, Appl
42	42.4	6.1	270	3 US-08-897-126-30	Sequence 30, Appl
43	42.2	6.1	5844	4 US-10-204-708-89	Sequence 89, Appl
44	42.2	6.1	6866	4 US-10-204-708-20	Sequence 20, Appl
45	42	6.0	357	4 US-09-621-976-16058	Sequence 16058, A

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpc-F18
; US-08-232-463-14
Query Match 12.3%, Score 85.6, DB 1, Length 7218;

Best Local Similarity 6.4%; Pred. No. 1.4e-12;
Matches 25; Conservative 232; Mismatches 131; Indels 0; Gaps 0.

[illegible]

```

RESULT 2
US-09-331-581-3
Sequence 3, Application US/09331581
Patent No. 6130070
GENERAL INFORMATION:
APPLICANT: TOHDA, Hideki
APPLICANT: HAMA, Yuko
APPLICANT: KIMAGAI, Hiromi-cho
TITLE OF INVENTION: INDICATOR PROMOTER GENE AND SECRETORY SIGNAL GENE
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
FILE REFERENCE: 0059-1142-OPT
CURRENT FILING DATE: US/09/331,581
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/JP98/04929
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: JP 9-314608
EARLIER FILING DATE: 1997-10-31
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 7286
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3

```

Query Match	7.6%	Score 52.6;	DB 3;	Length 7286;
Best Local Similarity	56.7%	Pred. No. 0.00044;		
Matches	97;	Conservative	0;	Mismatches 74;
			Indels	0;
			Gaps	0;

QY	9	AACAAAAATTAAAGCAATGTCATGTCGTAGCAATAAATTATATGTGGAAACAATTAGTTA	68
DB	4559	AACATACACTGAGAAAGATGTCTATGTAGCTGAAATTAATGACGTCAAGACAAAAA	4618
QY	69	ACGCGAAAAAGAAAAAAAAGGTCACAAAAATGAAAAACAAATCAAACTGAATGAAAAATT	128
DB	4619	AAG	4678
QY	129	GGAGTCCAGATCCGAAAAACGAGCCGTTTATAGACTTAATTAAGCTTCT	179

Db 4679 TACCTTCTGAGCGGAAAGAACAGCCGATCCAGACATGATAGATACAT 4729

```

RESULT 3
US-09-331-581-14
; Sequence 14, Application US/09331581
; Patent No. 6130070
; GENERAL INFORMATION:
; APPLICANT: TOMODA, Hideki
; APPLICANT: HAMA, Yuko
; APPLICANT: KIMAGAI, Hiromichi
; TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
; TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
; TITLE OF INVENTION: HAVING THE SAME, AND USE THEREOF
; FILE REFERENCE: 0059-1142-0PCT
; CURRENT APPLICATION NUMBER: US/09/331,581
; EARLIER FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: PCT/JP98/04929
; EARLIER FILING DATE: 1998-10-30
; EARLIER APPLICATION NUMBER: JP 9-314608
; EARLIER FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 7938
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14

```

```

Query Match      7.6%; Score 52.6; DB 3; Length 7938;
Best Local Similarity 56.7%; Pred. No. 0.00045;
Matches 97; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY      9 AACCAAAATTTAGCAAGTCATGTTCTCGAGCAATAAATTAATAGTGGAAACAATTAAGTTA 68
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      5211 AACCAATACATGGAAGAAGTGTCTATGTAGCTGAAATTAATAAGAGCTACAGACAAAA 5270

QY      69 ACGCAAAAAAGCAAAAAAAGGTACAAAAATGAAAAAACAATTCAAACTGAATGAAAAATT 128
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      5271 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAG 5330

QY      129 GGAGTCGAGATCGGAAAAACGAGCCGCTTTTAGAGCTTAATTAAGCTTCT 179
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      5331 TACCTTCTGAGGCGGAAAAAGAACGACCGGATCCAGACATGATTAAGATCAT 5381

RESULT 4
US-09-640-173-53
; Sequence 53, Application US/09640173
; Patent No. 6613515
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Scolk, John A.
; TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
; TITLE OF INVENTION: METHODS OF USE THEREFOR
; FILE REFERENCE: 210121.484C2
; CURRENT APPLICATION NUMBER: US/09/640.173
; CURRENT FILING DATE: 2000-08-15
; NUMBER OF SEQ ID NOS: 196
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G

US-09-640-173-53

```

Query Match 7.3%; Score 50.6; DB 4; Length 396;
Best Local Similarity 49.8%; Pred. No. 0.00049;
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 448
DB 18 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 77
QY 449 CGTCTACTGTTTGTTCATTTTGTGCTTTCACGTTTACGCTGATCGAGTAT 508
DB 78 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 137
QY 509 TTGACTGGAATAATCCCTGCTTTTGTGCTTTCATATAATCGATGATAC 568
DB 138 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 197
QY 569 CTTTGTGCTTGAATGTTGTTTGTGACCTATGCGTTGGCTGTATTAATCTCA 627
DB 198 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 256

RESULT 5
US-09-713-550-53
; Sequence 53, Application US/09713550
; Patent No. 6617109
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Stolk, John A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.484C4
; CURRENT APPLICATION NUMBER: US/09/713,550
; CURRENT FILING DATE: 2000-11-14
; NUMBER OF SEQ ID NOS: 205
; SOFTWARE: FastSeq For Windows Version 3.0
; SEQ ID NO 53
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(396)
; OTHER INFORMATION: n = A,T,C or G
US-09-713-550-53

Query Match 7.3%; Score 50.6; DB 4; Length 396;
Best Local Similarity 49.8%; Pred. No. 0.00049;
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 448
DB 18 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 77
QY 449 CGTCTACTGTTTGTTCATTTTGTGCTTTCACGTTTACGCTGATCGAGTAT 508
DB 78 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 137
QY 509 TTGACTGGAATAATCCCTGCTTTTGTGCTTTCATATAATCGATGATAC 568
DB 138 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 197
QY 569 CTTTGTGCTTGAATGTTGTTTGTGACCTATGCGTTGGCTGTATTAATCTCA 627
DB 198 TTTTGTGATAGGTTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCG 256

RESULT 6
US-10-204-708-4
; Sequence 4, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: Olek, Alexander
; APPLICANT: PIRENBROCK, Christian

; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; TITLE OF INVENTION: By Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 4
; LENGTH: 10619
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-4

Query Match 7.3%; Score 50.6; DB 4; Length 10619;
Best Local Similarity 49.8%; Pred. No. 0.0017;
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 404 TTACTTTTCCCTATATTATAGTCTTAGAGTAACGATACCTGCTTACTGTTT 463
DB 9137 TTATTTATTTGTTATTTATATATATATATATATATATATATATATATAT 9196
QY 464 GTTCAATTTTGTGCTTTCACGTTTACGCTGCTATCGAGTATTAATCTCA 523
DB 9197 TTTGCTTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 9256
QY 524 CTTGCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 583
DB 9257 AGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 9316
QY 584 GTTGTGTTTGTGCTTATGCTGTTGCTGTTTATTAATCTCACTGATGTT 643
DB 9317 GTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGT 9376
QY 644 TTGAGATTTTGTGCTG 660
DB 9377 TGTATTTATTTGTTAGTG 9393

RESULT 7
US-10-204-708-63
; Sequence 63, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: Olek, Alexander
; APPLICANT: PIRENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01

NUMBER OF SEQ ID NOS: 98
SEQ ID NO: 63
LENGTH: 5562
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-63

Query Match 7.2%; Score 50.2; DB 4; Length 5562;
Best Local Similarity 48.1%; Pred. No. 0.0017; Mismatches 153; Indels 0; Gaps 0;
Matches 142; Conservative 0;

QY 390 TTTTGGCATGAGGTTACTTTTCCATATTATTAAGTCTTAGTAAGCATACCTGC 449
DB 1394 TATTTTATTAATTAATTTTTCGTTAGGAATTTTTCGATTAATTAAT 1453
QY 450 GTCCTACTGTTTGTTCATTTTGTGCTTTCACCGTTAGTGGCATGAGTAT 509
DB 1454 ATTTTATTTAGTTAGTTAGTTTATTTTATTTTATTTTATTTATTTTAT 1513
QY 510 TGACTGTGAAAATCCTTCGTTTGTGTTTGCATATAAATCGATGATCTACC 569
DB 1514 TGTTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1573
QY 570 TTTTGTGCTTGTGATTTTGTGTTTGTGAGCCTAAGCGTTGCGTTTATTA 629
DB 1574 TTTTGTGTTTGAAGAAGTTATTAATTTTATTTTATTTTATTTTATTTT 1633
QY 630 TTCATGTGTGATTTTGAATTTTGTGATGACTGTGGTTCTTGTGGCTAT 684
DB 1634 GATAGCTTATTTTGAATTTTGTGATAATTAATTAATTAATTAATTTAT 1688

RESULT 8

US-08-487-826B-13/c
Sequence 13, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim. Kim L.
APPLICANT: Chelnis, Chean
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:

LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 7.1%; Score 49.4; DB 2; Length 19124;
Best Local Similarity 47.9%; Pred. No. 0.0042; Mismatches 186; Indels 1; Gaps 1;
Matches 172; Conservative 0;

QY 318 GCATGATCTGTGCGCATTTCTCTTTTCGCTGGAATAATGCCCTAATGCTTC 377
DB 15976 GCATGATTAATTAATTTATTTTATTTTATTTAATTAATTTTATTTTAT 15917
QY 378 GATTCGAGGTTTGTGCTATGGTTACTTTTCCATATTTATTAAGTCTAGGT 437
DB 15916 TTCAATTAATTTTATTTTATTTTATTTAATTAATTTTATTTATTTAT 15857
QY 438 AACGATACCTGCGCTTACTGTTTGTTCATTTGTGCTTTCACCGTTAGTGGCT 497
DB 15856 TTATTTAATTAATTTTATTTTATTTTATTTTATTTTATTTAATTTATTT 15797
QY 498 GATCGAGATTTGACTGTGAAAATCCTTCGTTTGTGTTTGTTCATATAAATCG 557
DB 15796 TATGATTAATTTTATTTTATTTTATTTTAAATTTTATTTTATTAATTAATTT 15737
QY 558 G-ATTGATCACTTTTGTGCTTGTGATTTTGTGTTTGTGAGCCATGCGTTGCTG 616
DB 15736 TTATTTAATTAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 15677
QY 617 TTATTAATCACTGCTTCAAGTGTGATTTTGTGATTTTGTGATGACTGTGGTTCTT 675
DB 15676 AATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTATAT 15618

RESULT 9

US-09-014-969-14/c
Sequence 14, Application US/09014969
Patent No. 5965397

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Racie, Lisa A.
APPLICANT: Treacy, Maurice
APPLICANT: Spaulding, Vikki
APPLICANT: Agostino, Michael J.
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
TITLE OF INVENTION: ENCODING THEM
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/014,969
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,323

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 2447 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-09-014-969-14

Query Match 6.9%; Score 47.8; DB 2; Length 2447;
Best Local Similarity 47.2%; Pred. No. 0.0051;
Matches 142; Conservative 1; Mismatches 158; Indels 0; Gaps 0;

QY 380 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 449
DB 2447 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 2388
QY 450 GCTTACGTTTGTGCTATTTGTGCTTCAACCGTTAGCGCTGATCGAGTAT 509
DB 2387 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 2328
QY 510 TGAAGTGAATAATCCTGCTTTTGTGCTTCAATATAACGATGATCACC 569
DB 2327 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 2268
QY 570 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 629
DB 2267 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 2208
QY 630 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 689
DB 2207 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACATACCTGC 2148
QY 690 G 690
DB 2147 G 2147

RESULT 10

US-10-204-708-29

Sequence 29, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPERROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
PRIOR FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 29
LENGTH: 5666
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-29

Query Match 6.8%; Score 47.4; DB 4; Length 5666;
Best Local Similarity 49.2%; Pred. No. 0.0088;
Matches 153; Conservative 0; Mismatches 156; Indels 2; Gaps 1;

QY 384 GAAGTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACAT 443
DB 3358 GATAGTATGCTATGATGCTGCTGCTATTAATTAATTAATTAATTAATTAAT 3417
QY 444 ACCGCGCTTACGCTTTTGTGCTATTTGTGCTTCAACCGTTAGCGCTGATCGG 503
DB 3418 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACAT 3477
QY 504 AGTATTTGCTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACAT 561
DB 3478 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACAT 3537
QY 562 GATCCTTGTGCTTGTGCTATTTGTGCTTCAACCGTTAGCGCTGATCGG 621
DB 3538 GATCCTTGTGCTTGTGCTATTTGTGCTTCAACCGTTAGCGCTGATCGG 3597
QY 622 ACTTACGCTTATGCTGCTATTTGTGCTTCAACCGTTAGCGCTGATCGG 681
DB 3598 TGTGCTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACAT 3657
QY 682 TATAGTGTGA 692
DB 3658 TATAGTGTGA 3668

RESULT 11

US-10-204-708-72

Sequence 72, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPERROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
PRIOR FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 72
LENGTH: 8607
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-72

Query Match 6.8%; Score 47.4; DB 4; Length 8607;
Best Local Similarity 46.7%; Pred. No. 0.01; Indels 0; Gaps 0;
Matches 150; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 374 TCTGATTCGAAGGCTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTT 433
DB 6966 TATGCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 7025
QY 434 AGGTACGATACCGCTTCTTACTGTTTGTGCTATTTGTGCTTCAACCGTTAGT 493
DB 7026 TTTTGTGCTAAGGCTTACTTTTCCCTAATTTATAGTCTTAGAGTAACAT 7085

QY 494 CGGTGATGAGATTTGACTGTGAAAAATCCCTGTTTGGTTTGTTCATATA 553
DB 7086 TTTTGTATTTTGTGTTTATATTTTATAGTTAGTTGTTTGGAGAA 7145
QY 554 ATCGGATGATGACCTTTGTGCTTGAAGTTTGTGAGCCATGCTGTGGC 613
DB 7146 ATAAAGAGTTTATGATGTTGTGTTGTTTATTTAGTTTATAGAGTTT 7205
QY 614 TTGTATACCTTACGTTGATGTGATTTTGAAGTTTGTGATGACTGTGGTCT 673
DB 7206 AGAATTAGTATTTTGTGATGTTAGTATTTGACGTGAAAGGATATGATTTTA 7265
QY 674 TTGGTGCTATAGTTGTTAA 694
DB 7266 AGTATGTTTGTGTTAA 7286

RESULT 12
US-10-708-79
Sequence 79, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
PRIOR FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 79
LENGTH: 8961
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
FEATURE: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (3866)
OTHER INFORMATION: n is a or g or c or t
US-10-204-708-79

Query Match 6.8%; Score 47; DB 4; Length 8961;
Best Local Similarity 48.9%; Pred. No. 0.013;
Matches 155; Conservative 0; Mismatches 160; Indels 2; Gaps 1;
QY 341 TCTTTCTTCGCTGAAAAATGCCATATGTTCTCGATTTGGAAGTTTGTGCTAT 400
DB 6013 TATTGTTGATGGGGAAATGTTGTTATTTTAAATGATGTTGATGTTATTT 6072
QY 401 GGGTACTTTTCCCATATTTTATAGTTCTAGTAACGATCCGCTCTACTGT 460
DB 6073 TAAGTACATGTTTTTTTTTGTGTTTATAGTTTAAAGTTTATGTTATTTAT 6132
QY 461 TTGTGATTTGTGTGCTTACCGTTAGTCGATCGAGATTTGACTGGA 520
DB 6133 ATTTTGTGATTTTATGTTTATGTTAATTTTATATATGTTATTTTAT 6190
QY 521 AATCCTTGTGTTTGTGTTTGTTCATATAATCGATGATGATCCTTTGTGCTT 580
DB 6191 TTTTGTGTTTATGTTTATGTTTATTTTGTGAGTTGTTGTTTGTGTTT 6250

QY 581 GATGTTGTTTGTGACCTATGCGTTGTGCTGTTATATCAGTCATGTCG 640
DB 6251 ATTTAGTATTTATTTTGTGTTTGTGTTTATGTTTATGAGAGTTGAGTGT 6310
QY 641 ATTTGAGATTTGTGA 657
DB 6311 TGGGAGATGTTTAT 6327

RESULT 13
US-10-204-708-45
Sequence 45, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204,708
PRIOR FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 45
LENGTH: 19233
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
FEATURE: chemically treated genomic DNA (Homo sapiens)
NAME/KEY: unsure
LOCATION: (76, 178, 179, 273, 586, 648, 651, 920, 1014, 1173, 1197, 1228)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (1264, 1267, 1559, 1607, 1677, 1703, 1742, 1848, 1850, 2033)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (2168, 2254, 2257, 2280, 2291, 2723, 3018, 3080, 4205, 4209)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (4216, 4225, 4239, 4246, 4254, 4265, 4270, 4274, 4288, 4294...4295)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (4298, 4363, 4610, 4627, 4675, 4693, 4698, 4738, 4781)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (4787, 4825, 4829, 4838, 4855, 5462, 5494, 5496, 5500, 5674)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (5690, 5697, 5705, 5715, 5890, 5910, 5913, 6027, 6080, 6109)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (6244, 6288, 6776, 6873, 6891, 6980, 7028, 7656, 7876, 8030)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure

LOCATION: (8059, 8075, 8632, 8636, 9700, 9717, 9815, 9818, 9867, 9879)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (10005, 10075, 10150, 10159, 10165, 10356, 11123, 11166, 11360)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (11703, 11933, 12050, 12340, 12981, 13564, 13764, 13765, 13787)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (13846, 14168, 16566, 17006, 17217, 17650, 17653, 17656, 17898)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (18130, 18143, 18151, 18188, 18498, 18549, 18936, 18946, 19165)
OTHER INFORMATION: n is a or g or c or t
FEATURE:
NAME/KEY: unsure
LOCATION: (19200)
OTHER INFORMATION: n is a or g or c or t
US-10-204-708-45

Query Match 6.7%; Score 46.8; DB 4; Length 19233;
Best Local Similarity 48.5%; Pred. No. 0.02;
Matches 129; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 360 AATGCGCTTAAGTCTGCATTCGAGGTTTGTGCTAATGGCTTACTTTTCCCTAT 419
DB 11391 AATTATATGTTTGTGAGATATGAAATATTTTATTAAGCTAATATTTTGT 11450
QY 420 AATTATATGTTTGTGAGATATGAAATATTTTATTAAGCTTACTTTTGTGCTG 479
DB 11451 TTTTATATGTTTGTGAGATATGAAATATTTTATTAAGCTTACTTTTGTG 11510
QY 480 TTTTACCGTTTGTGCTGATCGAGATATTTGACTGTGAAATATTTTGTGCTG 539
DB 11511 AGTTTGTGTTTATTAAGTTTATTAATTAATGAATTAATATTTTGTGCT 11570
QY 540 TTTGTTTCAATTAATCGAATGATCTACCTTTTGTGCTTGAATGTTTGTGAGCC 599
DB 11571 TTGCTTTTATTAAGTTTATTTTATTTTATTAAGGTTAGGTTTGTGTTT 11630
QY 600 TATGCTGTGCTGCTTGTATTAAT 625
DB 11631 GGAGTGTAGCATATATGTTATGTTT 11656

US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEPLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOMLOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463

FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE: EP 91 114 300.6
APPLICATION NUMBER: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZapc-F18
US-08-232-463-14

Query Match 6.6%; Score 46; DB 1; Length 7218;
Best Local Similarity 5.6%; Pred. No. 0.022;
Matches 10; Conservative 115; Mismatches 55; Indels 0; Gaps 0;

QY 1 GACAAACAACAAATTAAGCAAGTCTGCTAGCAATTAATTAAGTGGAAACA 60
DB 1231 RRR 1172
QY 61 TTAAGTACGCAAGAAAGAAAAAGTCAAAATGAATGAACAAATCAACTGAAT 120
DB 1171 RRR 1112
QY 121 GAAATTTGAGTCCAGATCGAAGAAACGAGCCGTTTATAGCTTAATAGCTTCC 180
DB 1111 RRR 1052

US-09-007-005-17/c
Sequence 17, Application US/09007005B
Patent No. 6258558
GENERAL INFORMATION:
APPLICANT: Szoestak, Jack W.
APPLICANT: Roberts, Richard W.
APPLICANT: Liu, Rife
TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
FILE REFERENCE: 00786/350003
CURRENT APPLICATION NUMBER: US/09/007,005B
CURRENT FILING DATE: 1998-01-14
EARLIER APPLICATION NUMBER: 60/035,963
EARLIER FILING DATE: 1997-01-27
EARLIER APPLICATION NUMBER: 60/064,491
EARLIER FILING DATE: 1997-11-06
NUMBER OF SEQ ID NOS: 33
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 289
TYPE: RNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Translation template
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(289)
OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 6.6%; Score 45.6; DB 3; Length 289;
Best Local Similarity 10.4%; Pred. No. 0.0085;
Matches 27; Conservative 105; Mismatches 128; Indels 0; Gaps 0;

```
QY 175 TTCCCTCATTTGCTCTCTCTGAGTTTATTTCTTCTCCGAGTCTGACTCAGTAC 234
    ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 275 TTTT TTTT TTTT TTTT TTAAGCYCYCAVAVGATYVAYACVGYCYAGVCYTGYSYN 216
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 235 TCTCACTCTCCGCGCCTTAACTTACGTTCTCCGTTACTCTGTAAGTTTCTGCC 294
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 215 YNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 156
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 295 TTRAGACCTCCGATCGCTCACCAGCATCTGTGCTGATTTCTTTCTTGCT 354
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 YNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 96
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 355 GGAATAATGCGCTTAATGTTCTGATTTGGAAGTTTGTGCTAAGGTTACTTTTTC 414
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 95 YNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYNYNYSYN 36
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 415 CCTATTTTATAGTTCTTA 434
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 35 YTVAVAYATYATYATYATA 16
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

Search completed: March 30, 2004, 03:14:52
Job time : 56.2914 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 13:41:30 ; Search time 242.442 Seconds
(without alignments)
12178.168 Million cell updates/sec

Title: US-10-030-829-1_COPY_1_695

Sequence: 1 gacaacaacaataa.....ggtgctatagtgtaa 695

Scoring table:

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
1000	1000	1000

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	695	100.0	3275	4	AAE23373	AAE23373 Genomic E
2	68.8	9.9	6668	6	ABL33697	ABL33697 Human Imm
3	68.6	9.9	556	5	ABV40163	ABV40163 Human pirc
4	68.6	9.9	556	5	ABV40063	ABV40063 Human pirc
5	68.6	9.9	556	5	ABV42105	ABV42105 Human pirc
6	68.6	9.9	556	5	ABV43601	ABV43601 Human pirc
7	66.6	9.6	9539	4	AA545347	AA545347 Chemical
8	66.6	9.6	9539	6	ABK28180	ABK28180 DNA trans
9	66	9.5	6644	2	AAK33181	AAK33181 Base seq
10	66	9.5	7372	2	AAK33182	AAK33182 Base seq
11	66	9.5	7797	2	AAK33180	AAK33180 Cowpo v
12	66	9.5	7996	2	AAK33184	AAK33184 Base seq
13	66	9.5	14006	6	ABL33958	ABL33958 Human imm
14	65.8	9.5	556	6	ABO36997	ABO36997 Oligonuc
15	65.8	9.5	556	6	ABO36996	ABO36996 Oligonuc
16	65.6	9.4	113515	6	ABL33174	ABL33174 Human Imm
17	65.4	9.4	6079	4	ABL34921	ABL34921 Human Imm
18	65.2	9.4	887	4	AAI94064	AAI94064 Human neu
19	65.2	9.4	887	7	ABT42734	ABT42734 Human neu
20	64.8	9.3	598	6	ABO52555	ABO52555 Oligonuc
21	64.8	9.3	598	6	ABO52554	ABO52554 Oligonuc
22	64.6	9.3	6668	6	ABL3696	ABL3696 Human Imm
23	64.4	9.3	850	6	ABQ48654	ABQ48654 Oligonuc

C	24	64.4	9.3	850	6	ABO48655	Abq948655	Oligonucleotide
	25	64.4	9.3	17934	6	AB133719	Ab133719	Human lymphocyte
	26	62.8	9.0	529	7	ABX52180	ABX52180	Bovine ES cell
C	27	62.8	9.0	629	7	ABT21705	ABT21705	Bovine calf thymus
	28	62.8	9.0	700	7	ACD92384	ACD92384	Human colon
	29	62.8	9.0	34769	4	AA546775	AA546775	Tumour suppressor gene
C	30	62.6	9.0	578	6	ABO52047	ABO52047	Oligonucleotide
	31	62.6	9.0	578	6	ABO52046	ABO52046	Oligonucleotide
C	32	62.6	9.0	875	4	AA195044	AA195044	Human neutrophil
	33	62.6	9.0	34769	4	AA546774	AA546774	Tumour suppressor gene
C	34	62.4	9.0	693	6	ABQ32105	ABQ32105	Oligonucleotide
	35	62.4	9.0	693	6	ABQ32104	ABQ32104	Oligonucleotide
	36	62.4	9.0	1200	6	ABO39210	ABQ39210	Oligonucleotide
C	37	62.4	9.0	1200	6	ABO39211	ABQ39211	Oligonucleotide
	38	62.4	9.0	8079	6	AB192213	AB192213	Chemical synthesis
	39	62.4	9.0	23659	6	ABQ66981	ABQ66981	Human angiotensin
	40	62.2	8.9	2357	6	ABBN0315	ABBN0315	Human cholesteryl ester
	41	62.2	8.9	6334	6	AB133212	AB133212	Human lymphocyte
	42	62	8.9	1179	6	ABO49604	ABQ49604	Oligonucleotide
C	43	62	8.9	1179	6	ABO49605	ABQ49605	Oligonucleotide
	44	62	8.9	8404	9	AD554216	AD554216	Pre-treated cells
	45	62	8.9	8404	9	AD584152	AD584152	Human lymphocyte

ALIGNMENTS

RESULT 1

ID AAF25373 standard; DNA; 3275 BP

AC AAF25373;

DT 15-MAY-2001 (first entry)

Genomic sequence of the Arabidopsis SGS3 gene

KM SGG3 gene; post-transcriptional inactivation; RNA degradation;
KM viral resistance; resistance; fatty acid content; protein content; ss
KM viral resistance; resistance; fatty acid content; protein content; ss

OS Arabidopsis thaliana

PH	Key	Location/Qualifiers
FT	exon	656..1658
FT		/*tag= a
FT	intron	1659..1731
FT		/*tag= b
FT	exon	1732..2023
FT		/*tag= c
FT	intron	2024..2134
FT		/*tag= d
FT	exon	2135..2379
FT		/*tag= e
FT	intron	2380..2481
FT		/*tag= f
FT	exon	2482..2648
FT		/*tag= g
FT	intron	2649..2738
FT		/*tag= h
FT	exon	2739..2949
FT		/*tag= i
XX		
PN	WO200105951-A2.	
XX		
PD	25-JAN-2001.	
XX		
PF	13-JUL-2000; 2000WO-FR002052.	
XX		
PR	16-JUL-1999; 99FR-00009417.	
PR	26-JAN-2000; 2000FR-00001006.	
XX		
PA	(AVET) AVENTIS CROPSHIELD SA.	
EN	(INRG) INST NAT RECH AGRONOMIQUE.	

XX Beclin C, Elmayer T, Vaucheret H;
XX WPI; 2001-159529/16.
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
XX resistance in plants and, when inhibited, for increasing transgene
XX expression.
XX Claim 1; Page 31-32; 36pp; French.
XX The present sequence represents the genomic sequence of the Arabidopsis
XX thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
XX inactivation (degradation of RNA) and for resistance to viruses.
XX Overexpression of SGS3 results in plants with increased resistance to
XX viruses, while inactivation of SGS3 in transgenic plants (e.g. by
XX expressing antisense RNA, by mutation or by homologous recombination)
XX increases the level of the transgene product. This product may e.g.
XX impart resistance (to herbicide, insects or pathogens), alter contents of
XX essential fatty acids or proteins, or is pharmaceutically active, e.g. an
XX immunoglobulin or interferon
SQ Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;
Query Match 100.0%; Score 695; DB 4; Length 3275;
Best Local Similarity 100.0%; Pred. No. 3.8e-145;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACAAACAACAAATTAAGCAAGTTCATGTCGATGCAATTAATTAATAGTGGACAA 60
DB 1 GACAAACAACAAATTAAGCAAGTTCATGTCGATGCAATTAATTAATAGTGGACAA 60
QY 61 TTAAGTAAAGCAAAAGGAAAAAGGTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 120
DB 61 TTAAGTAAAGCAAAAGGAAAAAGGTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 120
QY 121 GAAAAATTGGATCGAATCGAAAAAGAGCGGTTTAAAGCTTAATAGTCTCC 180
DB 121 GAAAAATTGGATCGAATCGAAAAAGAGCGGTTTAAAGCTTAATAGTCTCC 180
QY 181 ATTGTCCTCTTCGTCGATTTATTTCTCTCGAGATCCTGACTACTCTC 240
DB 181 ATTGTCCTCTTCGTCGATTTATTTCTCTCGAGATCCTGACTACTCTC 240
QY 241 TCTCCGCGCTTAAACCTTACGTTCCGCTGTTTACTGTAAGTTTCCCTTAAG 300
DB 241 TCTCCGCGCTTAAACCTTACGTTCCGCTGTTTACTGTAAGTTTCCCTTAAG 300
QY 301 CCTCGATGCGCTACCGGATGATTCGTCGATTTCTCTTCTTCTCGTGGAAAA 360
DB 301 CCTCGATGCGCTACCGGATGATTCGTCGATTTCTCTTCTTCTCGTGGAAAA 360
QY 361 ATTGCGCTAATGTCGATTTGAAAGTTTGTGCTAATGAGTTACTTTTCCCTATA 420
DB 361 ATTGCGCTAATGTCGATTTGAAAGTTTGTGCTAATGAGTTACTTTTCCCTATA 420
QY 421 TTTTATAGTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 480
DB 421 TTTTATAGTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 480
QY 481 TTTTATAGTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 540
DB 481 TTTTATAGTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 540
QY 541 TTTTATAGTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 600
DB 541 TTTTATAGTCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 600
QY 601 ATGCGTGTGCTGTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 660
DB 601 ATGCGTGTGCTGTTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 660
QY 661 ACTGTCGCTTCTTGTGCTAATAGTGTGAAAA 695

DB 661 ACTGTCGCTTCTTGTGCTAATAGTGTGAAAA 695
RESULT 2
ID ABL33697 standard; DNA; 6668 BP.
XX ABL33697;
XX 26-MAR-2002 (first entry)
XX Human immune system associated gene SEQ ID NO: 1670.
XX Human, immune system disease; cytosine methylation; antiaesthetic;
XX antiarteriosclerotic; antianemic; cytosolic; noctropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antiparasitic;
XX antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX Homo sapiens.
XX WO200200928-A2.
XX 03-JAN-2002.
XX 02-JUL-2001; 2001WO-EP007537.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPIG-) EPIGENOMICS AG.
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2002-130909/17.
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/necrotic bowel
XX diseases. The present sequence is a gene of the invention
SQ Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 0 U; 1 Other;
Query Match 9.9%; Score 68.8; DB 6; Length 6668;
Best Local Similarity 45.8%; Pred. No. 1.3e-05;
Matches 238; Conservative 0; Mismatches 282; Indels 0; Gaps 0;
QY 156 GTTTAGAGCTTAAGCTTCCATTTGCTCTCTCGAGTATTTCTTCTC 215
DB 2742 GTTTAGAGCTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2801
QY 216 CGAGTCTGACTCACTCACTCACTCGGCGCTTAAAGTTCGTCGCTT 275
DB 2802 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2861
QY 276 ACTGTAAGTTTCTGCTTAAAGCTCCGATCCCTCAGCATATTCGTCG 335
DB 2862 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 2921

XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 8100; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

Query Match 9.9%; Score 68.6; DB 5; Length 556; Best Local Similarity 48.3%; Pred. No. 8.6e-06; Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 265 CTCGCGTTACTGCTGAAGTTTCTGCTTAGACCTCCGATCGCCATCCGATGCA 324
DB CCCCCCTCCCCCTTTTATATATTTTATTTTCTCCCCCTCCCATTAATA 427

QY 325 TTCTGCTGATTTCTCTTTTCTGCTGAAAAATGCTTAATGTTCTGATTTG 384
DB TTTTCTTTTCTTTTCTTTTCTTTTATAAAAAAANCCCCCTTTTCTTTT 367

QY 426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 367

QY 385 AAGTTTGTGCTAGGTTACTTTTCTCCATAATTTAATGTTCTAGTAAGATA 444
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307

QY 366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307

QY 445 CTCGCGTTACTGTTTGTGCTATTTGTTGCTTACGCTTAGCGGATGCA 504
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247

QY 306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247

QY 505 GATTTGACTGAAAAATCTGTTTGTGTTTGTGTTTCTAATAATCGATGAT 564
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187

QY 246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187

QY 565 CTACCTTTGTGCTTGAATGTTTGTGAGCCATAGCGTTGTGCTTGAAT 624
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127

QY 186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127

QY 625 TCACGTTCAATGCTGATTTGAGATTTT 653
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 98

RESULT 5
ABV42105/c
ID ABV42105 standard; cDNA; 556 BP.
XX ABV42105;
AC
XX
DT 16-SBP-2002 (first entry)
DE Human prostate expression marker cDNA 42096.
XX
XX Human prostate expression marker cDNA 42096.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX phamacogenomic marker; gene; ss.
XX Homo sapiens.
OS
XX
XX WO200160860-A2.
XX
XX PD 23-AUG-2001.

XX 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
XX
XX 16-MAR-2000; 2000US-0189862P.
XX
XX 25-MAY-2000; 2000US-0207454P.
XX
XX 09-JUN-2000; 2000US-0213131P.
XX
XX 18-JUL-2000; 2000US-0219007P.
XX
XX 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
XX Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 8441; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

Query Match 9.9%; Score 68.6; DB 5; Length 556; Best Local Similarity 48.3%; Pred. No. 8.6e-06; Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 265 CTCGCGTTACTGCTGAAGTTTCTGCTTAGACCTCCGATCGCCATCCGATGCA 324
DB CCCCCCTCCCCCTTTTATATATTTTATTTTCTCCCCCTCCCATTAATA 427

QY 325 TTCTGCTGATTTCTCTTTTCTGCTGAAAAATGCTTAATGTTCTGATTTG 384
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 367

QY 426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 367

QY 385 AAGTTTGTGCTAGGTTACTTTTCTCCATAATTTAATGTTCTAGTAAGATA 444
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307

QY 366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 307

QY 445 CTCGCGTTACTGTTTGTGCTATTTGTTGCTTACGCTTAGCGGATGCA 504
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247

QY 306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 247

QY 505 GATTTGACTGAAAAATCTGTTTGTGTTTGTGTTTCTAATAATCGATGAT 564
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187

QY 246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 187

QY 565 CTACCTTTGTGCTTGAATGTTTGTGAGCCATAGCGTTGTGCTTGAAT 624
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127

QY 186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 127

QY 625 TCACGTTCAATGCTGATTTGAGATTTT 653
DB TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 98

RESULT 6
ABV43601/c
ID ABV43601 standard; cDNA; 556 BP.

[illegible]

Dd 246 TTTTTCCTTCCCTTTT 187

Oy 565 CTACCTTTTGCTGTGANGTTGGTTTTTGAGCCTAAGCGTTGGCTGTATAACT 624
| | | | |
Db 186 TT 127

Oy 625 TCACGTCAGTCGTGATTTTGAGATT 653
| | | | |
Db 126 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 98

RESULT 7
ID AAS45347 standard; DNA; 9539 BP.
XX AAS45347;
XX AC
XX AS45347;
Dd 18-DEC-2001 (first entry)
XX Chemically pretreated complementary DNA associated with cell cycle #26.
XX DE
XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiatherosclerotic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
XX PCR primer.
XX
OS Homo sapiens.
PN WO200168911-A2.
XX
PD 20-SEP-2001.
XX
PF 15-MAR-2001; 2001WO-BP002945.
XX
PR 15-MAR-2000; 2000DE-01013847,
PR 06-APR-2000; 2000DE-01013058,
PR 07-APR-2000; 2000DE-01019173,
PR 30-JUN-2000; 2000DE-01032529,
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIT-) EPIGENOMICS AG.
PI Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602751/68.
DR
XX Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
associated with cell cycle.

Claim 1; SEQ ID NO 52; 28bp; English.

PS Sequences AAS45296-AA45520 represent chemically pretreated genomic DNA
XX molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analysing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;

Query Match 9.6%; Score 66.6; DB 4; Length 9539;
 Best Local Similarity 46.9%; Pred. No. 4.4e-05;
 Matches 207; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 251 TTTAACTAGCTCTCGGTTTCTGTAAGTTTTCGCTTAGAGCCCTCCGATCG 310
 DB 202 TTTTCTTTTGTATTTTGTGTTTATTTTGTGTTTATTTTATTTTATTTTTC 261

QY 311 CCTCACCGATGATCTGCTGATTTCTTTTCTTGGCTGGAATAATGCCCTAA 370
 DB 262 GTTTTCTTTTGTGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTTT 321

QY 371 TGTTCGATTTGGAAGTTTGTGCTATGCTTATTTTCCCTATTTTATAGTT 430
 DB 322 TTAATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 381

QY 431 CTAGGTAAAGATACCTGCGCTTACTGTTTGTGCTATTTGTTGCTTACCGTT 490
 DB 382 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 441

QY 491 AGTCGCTATGAGATTTGACTGCTGGAATAATCCTGCTTTTGTGTTTGTTCATA 550
 DB 442 ATTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 501

QY 551 TAAATCGATGATACCTTTGCTTGTGATGTTTGTGTTTGTGAGCTATGCTGTT 610
 DB 502 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 561

QY 611 GCGTGTATTAACCTACGCTTACGTTACGTTGATTTTGAATTTGGAGAGCTGGTT 670
 DB 562 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 621

QY 671 TCTTGGTGGCTATAGTTGT 691
 DB 622 TTAATTAATATATTTT 642

RESULT 8
 ABRK28180
 ID ABRK28180 standard; DNA; 9539 BP.
 AC ABRK28180;
 XX 23-APR-2002 (first entry)
 DT
 XX
 DE DNA transcription associated complementary genomic DNA #27.
 XX
 XX DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
 KM PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
 KM single nucleotide polymorphism; adenine deaminase deficiency; cancer;
 KM immunological disorder; Sezary syndrome; haematological disorder; tuberculosis;
 KM psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
 KM neurodegenerative disorder; Waardenburg syndrome; myocardial infarction; hypertension; arthritis;
 KM myelodysplastic syndrome; congenital heart disease; gene therapy;
 KM angiodysplasia; congenital heart disease; HDR syndrome; gene therapy;
 KM polyglutamine disorder; solid tumour.
 OS
 XX
 OS Unidentified.
 XX
 PN WO200192565-A2.
 XX
 XX
 PD 06-DEC-2001.
 PD
 XX
 PF 06-APR-2001; 2001WO-EP003973.
 XX
 PR 06-APR-2000; 2000DE-01019058.
 PR
 PR 07-APR-2000; 2000DE-01019173.
 PR
 PR 30-JUN-2000; 2000DE-01032529.
 PR
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX
 PA (EPIC-) EPIGENOMICS AG.
 XX

PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-090046/12.
 DR
 XX New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological disorders,
 PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
 PT cancer.
 XX
 XX Claim 1; SEQ ID NO 54; 32bp; English.
 PS
 XX
 XX The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Rieger's syndrome,
 CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumors
 CC or cancer. Sequences ABRK28127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention. Note: The sequence
 CC data for this patent did not form part of the printed specification but
 CC was obtained in electronic format directly from the European Patent
 CC Office

Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;
 SQ

Query Match 9.6%; Score 66.6; DB 6; Length 9539;
 Best Local Similarity 46.9%; Pred. No. 4.4e-05;
 Matches 207; Conservative 0; Mismatches 234; Indels 0; Gaps 0;

QY 251 TTTAACTAGCTCTCGGTTTCTGTAAGTTTTCGCTTAGAGCCCTCCGATCG 310
 DB 202 TTTTCTTTTGTATTTTGTGTTTATTTTGTGTTTATTTTATTTTATTTTTC 261

QY 311 CCTCACCGATGATCTGCTGATTTCTTTTCTTGGCTGGAATAATGCCCTAA 370
 DB 262 GTTTTCTTTTGTGCTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTTT 321

QY 371 TGTTCGATTTGGAAGTTTGTGCTATGCTTATTTTCCCTATTTTATAGTT 430
 DB 322 TTAATGTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 381

QY 431 CTAGGTAAAGATACCTGCGCTTACTGTTTGTGCTATTTGTTGCTTACCGTT 490
 DB 382 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 441

QY 491 AGTCGCTATGAGATTTGACTGCTGGAATAATCCTGCTTTTGTGTTTGTTCATA 550
 DB 442 ATTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 501

QY 551 TAAATCGATGATACCTTTGCTTGTGATGTTTGTGTTTGTGAGCTATGCTGTT 610
 DB 502 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 561

QY 611 GCGTGTATTAACCTACGCTTACGTTACGTTGATTTTGAATTTGGAGAGCTGGTT 670
 DB 562 TTTTCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTATTT 621

QY 671 TCTTGGTGGCTATAGTTGT 691
 DB 622 TTAATTAATATATTTT 642

RESULT 9
AA33181/C
ID AAX33181 standard; DNA; 6644 BP.
XX
AC AAX33181;
XX
DT 25-JUN-1999 (first entry)
XX
DE Base sequence of the plasmid pRc-lres-bsr.
XX
KM Cowpox virus; bsr: viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KM autoimmune disease; graft rejection reaction; inflammation;
KM inflammatory disease; ss.
XX
OS Synthetic.
OS Cowpox virus.
XX
PN WO913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
XX New apoptosis-resistant virus-sensitive cell.
XX
PS Example 1; Page 38-41; 51pp; English.
XX
XX The present invention describes an apoptosis-resistant virus-sensitive
XX cell line into which an apoptosis resistance gene has been introduced.
XX The recombinant viruses generated are capable of expressing apoptosis-
XX associated genes. These can then be used in a variety of diseases for
XX which the induction of apoptosis by gene transfer, or where the
XX inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
XX are useful as vectors for gene therapy which can be applied to cancer
XX therapy for destroying cancer cells selectively, the treatment of
XX autoimmune diseases and graft rejection reaction, and apoptosis induction
XX therapy for inflammatory cells in inflammatory diseases. Prior arts have
XX encountered the problem where if an adenovirus vector capable of
XX expressing an apoptosis-associated gene is introduced into animal cells,
XX the cells producing the virus will be destroyed because the period of
XX time required to induce cell death by apoptosis is shorter than that
XX required to replicate and produce the virus, resulting in failure to
XX obtain a recombinant virus having the integrated apoptosis-associated
XX gene. In this invention an apoptosis-resistant 293 cell line (having an
XX apoptosis resistant gene introduced) is established and overcomes the
XX problem. The present sequence represents the base sequence of the plasmid
XX pRc-lres-bsr, which contains the cowpox virus bsr gene, and is used in an
XX example from the present invention
SQ Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 U; 0 Other;
Query Match 9.5%; Score 66; DB 2; Length 6644;
Best Local Similarity 45.4%; Pred. No. 5.6e-05;
Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0;
QY 157 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGTCAGTTATTTCTCTCC 216
DB 4235 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4176
QY 217 GAGAGCCGACCTACCTACCTCCTCCTCGGCGCTTAACTAGCTTCCGCTGTTA 276
DB 4175 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4116
QY 277 CTCGTAGATTCTTGCCCTTAGAGCCTCGATCGCTCAGCAGCATTCGTGCTCGA 336

DB 4115 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 4056
QY 337 TTTCTCTTTTCTTGCGCGAAAAATGCGCCTAAGTCTCGATTGCAAGTTTGTG 396
DB 4055 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3996
QY 397 CTATGGTACTTTTCTTCCCTAATTTTAAAGTCTTAGGTAAAGATACCTGCTTAC 456
DB 3995 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3936
QY 457 TCTTTTCTTCAATTTGTTGCTTTCACCGTTAGTGCCTGATGAGATTTGACTGT 516
DB 3935 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3876
QY 517 GAAAAATCCTTCGTTTCTTTGCTTTTGTCTCAATAAATGGAATGATCTTACTTGTG 576
DB 3875 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3816
QY 577 CTTCATGTTGTTTCTTTGAGCCCTAAGCGTGTGCGCTTGTATACTCAGCTTCAGT 636
DB 3815 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3756
QY 637 GTGATTTTGAGATTTTGTGTAAGTCTGTGCGTTCTTTGCT 678
DB 3755 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3714
RESULT 10
AA33182/C
ID AAX33182 standard; DNA; 7372 BP.
XX
AC AAX33182;
XX
DT 25-JUN-1999 (first entry)
XX
XX Base sequence of the plasmid pRc-Bcl-x1-bsr.
XX
XX
XX Cowpox virus; bsr: viral vector; expression; apoptosis; resistance; crmA;
XX bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;
XX autoimmune disease; graft rejection reaction; inflammation;
XX inflammatory disease; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN WO913073-A2.
XX
PD 18-MAR-1999.
XX
PF 07-SEP-1998; 98WO-JP004010.
XX
PR 08-SEP-1997; 97JP-00259235.
XX
PA (RPRG-) RPR GENCELL ASIA PACIFIC INC.
XX
PI Hamada H;
XX
DR WPI; 1999-243728/20.
XX
XX New apoptosis-resistant virus-sensitive cell.
XX
PS Example 2; Page 41-45; 51pp; English.
XX
XX The present invention describes an apoptosis-resistant virus-sensitive
XX cell line into which an apoptosis resistance gene has been introduced.
XX The recombinant viruses generated are capable of expressing apoptosis-
XX associated genes. These can then be used in a variety of diseases for
XX which the induction of apoptosis by gene transfer, or where the
XX inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
XX are useful as vectors for gene therapy which can be applied to cancer
XX therapy for destroying cancer cells selectively, the treatment of
XX autoimmune diseases and graft rejection reaction, and apoptosis induction

the therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where, if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the base sequence of the plasmid pRc-Bcl-1, which contains the human Bcl-1 gene, and is used in an example from the present invention.

Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 U; 0 Other;

Query Match	Score	DB 2	Length
9.5%	66		7372

Best Local Similarity 45.4%; Pred. No. 5.7e-05;

Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0

QY	157	TTTTAGACTTAATAGAGTTCCTGATTCGTCCTCTTCGGCAATTAATTTCTCTCC	216
Db	4963	TT	4904
QY	217	GGAGCTCGACTCACTACTCTCACTCCGAGCGCTTAACTCAAGTTCGCGTTTA	276
Db	4903	TT	4844
QY	277	CTCTGAAGTTTCGCTTAGAGCTCCGACGCGCTACCGCAGCAGCATCTGCGCTGA	336
Db	4843	TT	4784
QY	337	TTTCTCTTTTCTTCGCTGAAAAATTCCTAATGTCGATTCGAAAGTTTGTG	396
Db	4783	TT	4724
QY	397	CTAAGGTAAGTTTTCCTTAATTTAATAGTCTTAGAGATACCGGCTCTAC	456
Db	4723	TT	4664
QY	457	TGTTTTTGTCATTTGTGTGCTTCAACGTTAGTCGATCGAGTAATTAAGCTGT	516
Db	4663	TT	4604
QY	517	GAAAAATCCTGCTTTTGGTTTGGTTTCAATAAATCGATTCACCTTTGTG	576
Db	4603	TT	4544
QY	577	CTTTGATGTTGTTTTTGAGCCTAATGCGTTTGGCTGTATTAACCTCAGTTGAT	636
Db	4543	TT	4484
QY	637	GTGAGATTTGAGATTTGGTAGTGAAGGGTTCTTTGGT	678
Db	4483	TTTTTTTTTTTTTTTTTGCAATCGAAGGGGTTTAATGAT	4442

RESULT 11
AAx33180/c
ID AAX33180 standard; DNA; 7797 BP.

AC AAX33180;

DT 25-JUN-1999 (first entry)

DE Cowpox virus bsr full length gene sequence.

KW Cowpox virus; bsr, viral vector; expression; apoptosis; resistance; crmA;

KW bcl-2; bcl-x1; FLIP; survivin; IAP; ILP; adenovirus; cancer;

KW autoimmune disease; graft rejection reaction; inflammation;

KW Inflammatory disease; ss-
YY

OS Cowpox virus.

PN - WO9913073-A2.

PD 18-MAR-1999

PF 07-SEP-1998; 98WO-JP004010.

PR 08-SEP-1997; 97JP-00259235.

PA (RPRG-) RPR GENCELL ASIA PACIFIC INC

PI Hamada H; xv

DR WPI; 1999-243728/20
XX

PT New apoptosis-resistant virus-sensitive cell.

Example 1; Page 34-38; 51pp; English
PS
XX

The present invention describes an apoptosis-resistant virus-sensitive cell line into which an apoptosis resistance gene has been introduced. The recombinant viruses generated are capable of expressing apoptosis-associated genes. These can then be used in a variety of diseases for which the induction of apoptosis by gene transfer, or where the inhibition of harmful apoptosis, is therapeutic. The recombinant viruses are useful as vectors for gene therapy which can be applied to cancer therapy for destroying cancer cells selectively, the treatment of autoimmune diseases and graft rejection reaction, and apoptosis induction therapy for inflammatory cells in inflammatory diseases. Prior arts have encountered the problem where if an adenovirus vector capable of expressing an apoptosis-associated gene is introduced into animal cells, the cells producing the virus will be destroyed because the period of time required to induce cell death by apoptosis is shorter than that required to replicate and produce the virus, resulting in failure to obtain a recombinant virus having the integrated apoptosis-associated gene. In this invention an apoptosis-resistant 293 cell line (having an apoptosis resistant gene introduced) is established and overcomes the problem. The present sequence represents the cowpox virus bsr gene which is used in an example from the present invention.

SQ Sequence 7797 BP; 2542 A; 1760 C; 1656 G; 1839 T; 0 U; 0 Other

Query Match	9.5%	Score 66	DB 2	Length 7797
-------------	------	----------	------	-------------

Best Local Similarity 45.4%; Pred. No. 5.8e-05;

Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0

[illegible]

PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.

XX Claim 1; SEQ ID NO 1931, 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel
CC diseases. The present sequence is a gene of the invention

XX Sequence 14006 BP; 3278 A; 155 C; 3257 G; 7313 T; 0 U; 3 Other;

Query Match 9.5%; Score 66; DB 6; Length 14006;
Best Local Similarity 45.3%; Pred. No. 6.5e-05;
Matches 240; Conservative 0; Mismatches 290; Indels 0; Gaps 0;

QY 157 TTTTGGAGCTTAAAGCTTCCGATTTGCTCTCTCTGTCAGTTATTTCTTCCTCC 216
DB 1974 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2033
QY 217 GGAGTCTGACTACTACTCTGACTCTCGGGCTTTAACTTACGCTCTCCGCTGTTA 276
DB 2034 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2093
QY 277 CTCTGTAAGTTTTCGCTTAAAGCTTCCGATTTGCTCTCTGTCAGTTATTTCTTCCTCC 336
DB 2094 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2153
QY 337 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 396
DB 2154 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2213
QY 397 CTATGGATTTCTTTTCCCTATATTTATTTAGTTCTTAAAGTACCTGCTTAC 456
DB 2214 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2273
QY 457 TGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 516
DB 2274 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2333
QY 517 GAAAAATCTGCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 576
DB 2334 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2393
QY 577 CTGTGATGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 636
DB 2394 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2453
QY 637 GTGATTTTGAATTTGTAGTACTGTGGTTTCTTGTGCTATAG 686
DB 2454 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 2503

RESULT 14
ABO36997/c
ID ABO36997 standard; DNA; 556 BP.

XX ABO36997;
XX
XX 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 23588.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.

OS Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EP010074.

XX 01-SEP-2000; 2000DE-01043826.

XX 05-SEP-2000; 2000DE-01044543.

XX (EPIC-) EPICENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly for detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABO13410-
CC ABO44121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention

XX Sequence 556 BP; 505 A; 14 C; 13 G; 12 T; 0 U; 12 Other;

Query Match 9.5%; Score 65.8; DB 6; Length 556;
Best Local Similarity 49.3%; Pred. No. 3.6e-05;
Matches 172; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

QY 325 TTTGTCGCTGCAATTTCTCTTCTGCTGAAAAATGCCCCATATGTTCTGATTTG 384
DB 461 TTTGTCGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 402
QY 385 AAGGTTTGTGCTATGCGTTACTTTTCCCTATATTTATTTAGTTCTTAAAGTAA 444
DB 401 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 442
QY 445 CCGGCTTACTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 504
DB 341 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 282
QY 505 GATTTGACTGTGAAAAATCTGCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 564
DB 281 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 222
QY 565 CTACCTTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 624
DB 221 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 162
QY 625 TCACTGTACATGCTGATTTGTAGTACTGTGGGTTCT 673
DB 161 TTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 113

RESULT 15

ABQ36996

ID ABQ36996 standard; DNA; 556 BP.

AC ABQ36996;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 23587.

KM Human, cytosine methylation; 5'-CpG-3', uracil; cytosine, diagnosis;
KM drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
KM SNE; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP010074.

PR 01-SEP-2000; 2000DE-01043826.

PR 05-SEP-2000; 2000DE-01044543.

PA (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

DR WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful for
PT diagnosis and prognosis, comprises selective hybridization of amplicons
PT from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German

This invention describes a novel method for determining the degree of methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridization to both classes is determined from the label on the amplicon. From the ratio of labels hybridized to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNPs), and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. ABQ3410-ABQ3412 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention

SQ Sequence 556 BP; 12 A; 13 C; 14 G; 505 T; 0 U; 12 Other;

Query Match 9.5%; Score 65.8; DB 6; Length 556;

Best Local Similarity 49.3%; Pred. No. 3.6e-05;

Matches 172; Conservative 0; Mismatches 177; Indels 0; Gaps 0;

325 TTCTGTGCTCGATTCTCTTTTCTTGGCTGGAATAATGGCCCTAATGTTCTCGATTTCG 384

Db 96 TTTCCGTTTTTTTTTTTTTTTCGTTATTTTTTTTTT 155

385 AAGTTTGTGCTATGGCTACTTTTCCCTATATTATAGTCTTAGGTAACGATA 444

Dd 156 TTTTCTTTTTTTACGTTTTTCGTTTTTTTTTTTT 215

445 CCTGGCTTACTGTTTGTTCATTGGTGTGCTTTCACCGTTTAGTCGCTGATCGGA 504

Dd 216 TTTTTTTTTTTTTTTTTTTTTTTTCGGTTTTTTTTTTTTTAA 275

505 GTATTGACTGTGAAAAATCCTTCGTTTTTTGGTTTTGTTTCATATATAATCGATTGAT 564

[illegible]

565 CTACCTTTGTGCTTGAAGTTGTTTTTGAGCCTATGCGTTGTGGCTTTATACT 624

Dd 336 TTTTTTTTTTTTTTTTTTTTTTTTTTTTAAATTT 399

625 TCACGTTTCATGTGTGGAATTTTGAGATTTTGGTAGTGACTGTGGTTTCT 673

D 396 TTTTTTTTTTTTTTTTTTTT CGTTTTTT CGTTTTTT 444

Search completed: March 29, 2004, 17:14:35
Job time : 249.442 Secs

Job time : 249.442 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 14:52:55 ; Search time 2617.78 Seconds

(without alignments)
11507.244 Million cell updates/sec

Title: US-10-030-829-1_COPY_1_695

Perfect score: 695
Sequence: 1 gacaaacaacaataa.....gtcgcatagttgtaaa 695

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_stb: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_or: *
21: em_ov: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rtd: *
36: em_htg_mam: *
37: em_htg_vrl: *
38: em_sy: *
39: em_htg_hum: *
40: em_htg_mus: *
41: em_htg_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	695	100.0	3275	6 AX078760	AX078760 Sequence
2	695	100.0	81365	8 AB025633	AB025633 Arabidops
3	85.6	12.3	7218	6 166494	166494 Sequence 14
4	84	12.1	2162	8 BT002944	BT002944 Arabidops
5	83.4	12.0	184535	2 CENS05TCJ	AL355100 Homo sapi
6	81.2	11.7	349980	6 AX344566	AX344566 Sequence
7	72.8	10.5	1407	8 AJ592026	AJ592026 Arabidops
8	70	10.1	1453	8 AJ592978	AJ592978 Arabidops
9	70	10.1	66993	2 AC138074	AC138074 Homo sapi
10	69.6	10.0	1434	8 AJ592058	AJ592058 Arabidops
11	69.2	10.0	3364	14 TSP418778	AJ481778 Tomato sp
12	69	9.9	88932	2 AC022680	AC022680 Homo sapi
13	68.8	9.9	6666	6 AX346599	AX346599 Sequence
14	68.8	9.9	81120	2 AC022851	AC022851 Homo sapi
15	68.4	9.8	810	8 AJ592180	AJ592180 Arabidops
16	68.2	9.8	56857	2 AC021917	AC021917 Homo sapi
17	68	9.8	1184	8 AJ591982	AJ591982 Arabidops
18	67.6	9.7	151368	2 AC142664	AC142664 Macaca mu
19	67.4	9.7	156550	2 AC015830	AC015830 Homo sapi
20	67.4	9.7	196853	9 CENS01DUD	AL133240 Human chr
21	67	9.6	1228	8 AJ592059	AJ592059 Arabidops
22	67	9.6	60565	2 AC023852	AC023852 Homo sapi
23	67	9.6	127709	2 CENS05TCI	AL352983 Homo sapi
24	67	9.6	202083	2 AC023833	AC023833 Mus muscu
25	66.8	9.6	164288	9 AC128949	AC128949 Rattus no
26	66.6	9.6	749	9 HSA336745	AJ336745 Homo sapi
27	66.6	9.6	9539	6 AX277889	AJ277889 Sequence
28	66.6	9.6	9539	6 AX323566	AX323566 Sequence
29	66.4	9.6	83440	2 AC024285	AC024285 Homo sapi
30	66.4	9.6	194180	2 AC069140	AC069140 Homo sapi
31	66	9.5	6644	6 E23356	E23356 Virus vecto
32	66	9.5	7372	6 E23357	E23357 Virus vecto
33	66	9.5	7797	6 E23355	E23355 Virus vecto
34	66	9.5	7996	6 E23359	E23359 Virus vecto
35	66	9.5	14006	6 AX346860	AX346860 Sequence
36	66	9.5	56693	2 AC084077	AC084077 Homo sapi
37	66	9.5	143585	2 AC013349	AC013349 Homo sapi
38	65.8	9.5	64767	2 AC102701	AC102701 Mus muscu
39	65.8	9.5	273807	2 AC025421	AC025421 Homo sapi
40	65.8	9.5	289973	2 AC135678	AC135678 Rattus no
41	65.6	9.4	113515	6 AX347076	AX347076 Sequence
42	65.4	9.4	6079	6 AX345323	AX345323 Sequence
43	65.4	9.4	93791	2 AC138073	AC138073 Homo sapi
44	65.4	9.4	174384	2 AC009524	AC009524 Homo sapi
45	65.4	9.4	242885	2 AC097836	AC097836 Rattus no

ALIGNMENTS

RESULT 1
AX078760 3275 bp DNA linear PAT 22-FEB-2001
LOCUS Sequence 1 from Patent WO0105951.
DEFINITION AX078760
ACCESSION AX078760 GI:13158379
VERSION AX078760.1 GI:13158379
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidops.
REFERENCE
1 Beclin,C., Elmayan,T. and Vaucheret,H.
Novel sg3 plant gene and use thereof

JOURNAL Patent: WO 0105951-A 1 25-JAN-2001;
 AVENTIS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
 AGRONOMIQUE (FR)

FEATURES
 source Location/Qualifiers
 1..3275
 /organism="Arabidopsis thaliana"
 /mol_type="unassigned DNA"
 /db_xref="taxon:3702"
 primer_bind /note="p356AD"
 693..715
 primer_bind complement (2926..2952)
 /note="p356Y"

ORIGIN

Query Match 100.0%; Score 695; DB 6; Length 3275;
 Best Local Similarity 100.0%; Pred. No. 1.2e-130; Mismatches 0; Gaps 0;
 Matches 695; Conservative 0; Indels 0;

1 GACAAACAAACAAATTAAGCAAGTATGTTCTGACCAATTAATTAATGAGGAAACA 60
 1 GACAAACAAACAAATTAAGCAAGTATGTTCTGACCAATTAATTAATGAGGAAACA 60

61 TTAAAGTTAGGAAAAAGAAAAAGTACAAAAATGAAACAAATCAATCAAT 120
 61 TTAAAGTTAGGAAAAAGAAAAAGTACAAAAATGAAACAAATCAATCAAT 120

121 GAAATTTGAGTCCAGATCGAAAAACGAGCGTTTATAGACTTAATTAAGCTTCTC 180
 121 GAAATTTGAGTCCAGATCGAAAAACGAGCGTTTATAGACTTAATTAAGCTTCTC 180

181 ATTTGCTCTTCTTCTGCAATTAATTTCTTCTCCGAGTCTGACCTACATCTCCAC 240
 181 ATTTGCTCTTCTTCTGCAATTAATTTCTTCTCCGAGTCTGACCTACATCTCCAC 240

241 TCTCCGGGCTTAACTAGTTCCTGCTGCTTACTCTGTAAGTTTCTGCTTAAAG 300
 241 TCTCCGGGCTTAACTAGTTCCTGCTGCTTACTCTGTAAGTTTCTGCTTAAAG 300

301 CCTCCGATGCTCAACCGATGATCTGCTGCTGATTTCTTTTCTGCTGGAAGA 360
 301 CCTCCGATGCTCAACCGATGATCTGCTGCTGATTTCTTTTCTGCTGGAAGA 360

361 ATTCGCTTAATGTTCTGATTTGCAAGTTTGTGCAATGGGTACTTTTCCCTATA 420
 361 ATTCGCTTAATGTTCTGATTTGCAAGTTTGTGCAATGGGTACTTTTCCCTATA 420

421 TTTTATAGTTCTTAGTAACGATACCTGCTGCTTACTGTTTGTTCATTTTGTGCT 480
 421 TTTTATAGTTCTTAGTAACGATACCTGCTGCTTACTGTTTGTTCATTTTGTGCT 480

481 TTACACGTTTAACTGCTGATCGAGATTTGACTGTGAAAAATCCTGTTTTTGGTTT 540
 481 TTACACGTTTAACTGCTGATCGAGATTTGACTGTGAAAAATCCTGTTTTTGGTTT 540

541 TTGTTTCAATTAATGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 600
 541 TTGTTTCAATTAATGAGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 600

601 ATGCGTTTGGCTGTTTAACTTCAACGTTCAATGATGATTAATGATTAATGATTAATGATTAAT 660
 601 ATGCGTTTGGCTGTTTAACTTCAACGTTCAATGATGATTAATGATTAATGATTAATGATTAAT 660

661 ACTGTTGGGTTTCTTTGGTGGCTTAAAGTTTATAAA 695
 661 ACTGTTGGGTTTCTTTGGTGGCTTAAAGTTTATAAA 695

RESULT 2
 AB025633 81365 bp DNA linear PLN 27-DEC-2000
 LOCUS Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MOM1.
 DEFINITION
 ACCESSION AB025633 BA000015
 VERSION AB025633.2 GI:10178221

KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (sites)

REFERENCE
 AUTHORS Sato, S., Nakamura, Y., Kaneko, T., Katoh, T., Asanizu, E., Kotani, H.
 and Tabata, S.
 TITLE Structural analysis of Arabidopsis thaliana chromosome 5. X.
 JOURNAL DNA Res. 7 (1), 31-63 (2000)
 MEDLINE 20181125
 PUBMED 10718197

REFERENCE
 AUTHORS Nakamura, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research
 Institute, Department of Plant Gene Research, 1532-3, Yana,
 Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp,
 Tel: 81-438-52-3935, Fax: 81-438-52-3934)
 On Sep 15, 2000 this sequence version replaced gi:4589439.
 Address for correspondence: kaos@kazusa.or.jp
 For the latest information on annotation of this clone, please see
 http://www.kazusa.or.jp/kaos/cgi-bin/agd.graph.cgi?c=MOM1
 Genes with similarity to proteins in the databases are described in
 'product' or 'note' qualifiers. Genes that have no significant
 protein similarity are described as 'unknown protein'.
 The software programs used to predict genes include: Grail
 (Informatics Group, Oak Ridge National Laboratory,
 http://compbio.ornl.gov/Grail-1.3/),
 NERSCAN (Chris Burge, MIT, http://CBS-081.mit.edu/NERSCAN.html),
 NetGene2 (S.M. Hebsgaard, et al., CBR, Technical University of
 Denmark, http://www.cbs.dtu.dk/services/netgene2/) and
 SplicePredictor (Volker Brendel, Stanford University,
 http://gremlin.zool.iastate.edu/cgi-bin/sp.cgi).
 Genes encoding tRNAs are predicted by tRNAscan-SE
 (Sean Eddy, Washington University School of Medicine, St. Louis,
 http://genome.wustl.edu/eddy/tRNAscan-SE/).
 This sequence may not be the entire insert of this clone. It may be
 shorter because we remove overlaps between neighboring submissions.
 The 5' clone is K19M13 and the 3' clone is M011.

FEATURES
 source Location/Qualifiers
 1..81365
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /strain="Columbia"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="MOM1"
 /clone_id="Mitsui P1"
 join (3184..3403,3474..3595,3681..3723,3777..3898,
 4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)
 /note="gb|AA01580.1
 gene_id:MOM1.24
 similar to unknown protein"
 /codon_start=1
 /evidence=not experimental
 /protein_id="AA97251.1"
 /db_xref="GI:8809710"
 /translacion="MHVOPGTSVDSRAHKKVDTPPKHVOFLPISKTIVDDAONS
 YGSAFDPPSSNSPLSPVPEPSSSTSEGGDDPLFGIEDLQISGPRYGHETQACG
 YSINGTSCNFEWVCHLEDGSVNYIDSSRRSTNAKQPNYLVTAADDVLIATIEY
 QPLDDRKRGKRFAPKRYDLTVLIDVHTLTFEADCFVFEFMWPAIFICYLISGG
 ACEDPMOSNIKTLHTGHASYSKVSILAVGFVDIWEAATLSIKRGSYIKICISDLTIE
 KFSASTYTIPIRGOPARLVIIIGSDGSHSLRADGSPDLIGSRDEIVLTALFLKRW
 FFKANSFYHFRCSCHKSRGL"
 join (6317..6428,6508..6629,6714..6756,6841..6932,
 7121..7169,7296..7466,7550..7663,7785..7882,8112..8181)
 /note="gb|AA01580.1
 gene_id:MOM1.23
 similar to unknown protein"

CDS

CDS

FEATURES

source

/codon_start=1
/evidence=not experimental
/protein_id="BAA97250.1"
/db_xref="GI:8809709"
/translation="MOPSTSYSESPALNADETPISTKTVNPTKNIFFPDGDPAFL
PALNLDISGPIYGHLEQACGISINGTSCNFWCHLEDGSVNYIDGAKPYLVT
ADVDGLCLAEVQPLDIDRNRKGLVAVFANDNRKIACHPEMOSINDKTLGHASVY
SLAIGFVHWEATLSIEREGYTIKCNNDLITTEKFSATVKIPEKPAELVIGSD
GSEHCLVDNEMPDIRLCKEERVLRLVKKVVEOKL"
complement(join(8642..8705,8794..8863,8978..9387,
9584..10347))
/note="gene_id:MOML.22
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97249.1"
/db_xref="GI:8809708"
/translation="MSWMKKGSGMTAFDKORQKQIESEVEGDPPPVSTVNASF
GVRGLRLRNHSPSEKFSVLLPSPRPALTEKDCQNERGCCRKPTLILPVNS
HDLAFTLKEMNSWADNLIIRDVLSTEDPEMALATLKGVSGKEDDEPTIEGY
SDNRSEVYRPEKTVTSVYMAARSTPEDAKVYDLNDSGSSFLVNASNKEKPDII
SELDSTIQROASTIPEPMEBDLYLSHRDALKVMSASNSHSAQNAQORVHAA
KQSDKREBDMIAELKLAELAKKIIGITNKDNDIMKLDLHGLHATAVQLOERLW
IEGHFTVNRSVSPNRGRKNAALASAQEPFGRLEDEGMHCQRTSRELNSLOVITG
IGHSRGOASLPLAVKTFPEDNRYRFDETRGVITVPRKFRHS"
complement(12417..13424)
/note="contains similarity to unknown protein
gb|AA97018.1
gene_id:MOML.21"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97248.1"
/db_xref="GI:8809707"
/translation="MATDSQPNOKLTPLKTRIALTVISTMTDNOAGRPDGTNRFR
LFRPAPNPKPNVISTSDPVDSQDLFRLLTPVSGDKLPPVFFHGGGAFIS
PNAPYDNCRRPARKLPAYIVISNYLAPRHPAPYDQGFALKIENHGSILTA
NBLISRCFPAGDSNGNIANVAVIRICRPSRSTPAVIGLISIQPFGESEETAE
KQVAGPLVSPDRDTCWAKMGLNDRHEAVNADISGLDYPETMTVAVGFPL
KDMORSYEWMLKCGKATLIEYPMFHAFLPELPEAGLIMRIDPVDERAASIS
A"
join(14922..14981,15072..15167,15563..15666,15755..15891,
16096..16178)
/note="gene_id:MOML.20"
/codon_start=1
/evidence=not experimental
/product="50S ribosomal protein L24"
/protein_id="BAA97247.1"
/db_xref="GI:8809706"
/translation="MGWKAELKLRHKILRGDNVMIIRGDKETGTIRVIRSNR
VIEGKMLIKKHIKGPDHEGIFVYAPLHASVQVVDVTPGRPCVGVKGLJEDTK
VRVARGTSGSIIIPREILKIRAPRTAGKPDTEMEFVQTYDAKTKGMPUL"
join(16823..16952,17143..17320,17412..17493,17675..17851,
17943..18110,18199..18390)
/note="gene_id:MOML.19"
/codon_start=1
/evidence=not experimental
/product="26S proteasome, non-ATPase regulatory subunit"
/protein_id="BAA97246.1"
/db_xref="GI:8809705"
/translation="MERLQRIFGAGGLGHASDPSPTLDTSEQVYISLALLMKLKHG
RACVPMVGLMGEFVDEYVAVVDFAPQSGTGVSVAVDHFPTNLLDKLQKQ
RPEMVVGVYSHSGFGCWSGVDINTQSEALNORAVAVDPIOSVKKVVIDR
SINPOTIMGOEPRQTTNSGHLNKDSIQALLHGLNHYVAVSIAINRKNLEKMLIN
LHKKKTDLGLTLRFPDTHSKTNEQTVQEMSLAKTKAVQEEDELSPEKLAIVNWR
QDAKKHLEHVSIMNSNIVQTLGTMIDTVF"
complement(join(18965..19043,19240..19298,19419..19536,
19666..19733,19908..19961,20387..20530))
/note="contains similarity to unknown protein
emb|CAA22897.1
gene_id:MOML.18"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97245.1"

/db_xref="GI:8809704"
/translation="MDKMOAIEKMKLVGMEVEDEQQAADBSLSFMEDLNRCAL
TTRQRFYFALICLSAGLCTLLMLVFNFKVIGTTLNMAIGSTALIGOROV
TMDLDPARITATLALYIASITIALFCALYVKNKLLTLAIIILEFGLIWSLSTYPRR
TWSKIPMTCEPTEP"
join(22672..23634,23708..23999,24111..24355,24458..24624,
24715..24925)
/note="emb|CAB62356.1
gene_id:MOML.17
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97244.1"
/db_xref="GI:8809703"
/translation="MSGRAGPMSEKERNVGGYRPEVQIYQIAGTFLASSODNGEW
EYISKKNKRPNTSGKTVVSONSNPRAGGQOQGHGVSNGNNSGRNGNGC
IQANISGRGALSRYDDNNFVAPPVPVSRPLLEGMMQARGSQHTAVQEPVED
VDNAISEBNDSDALDSDDDVDVDSQSHSGRKONKMFKFFGLSDSIEQ
INBPQOMHPACONGGALIDVNLHPLHAHRTKGRVYKLRHLEAVELEKLOMG
ASVTPGEITVGOKGIGEDKQYEWPMVITIMTRLDKDDNDKTMGMNOELLEYF
DKYELRARSYGPQHRGMSVTFESSAGCYLEALRHELEAMEGLDRTANQKRM
PSCGVRQYFLTKQDLDLFGHSGQKTRLKELSYQEMVARELQISEDNQOLY
FKRLSKONKHAVALLESLEIMSEKLRTAEDNRYQRKTMQHEONREMDADRF
MDSIKQIHERDRAKEENFEMLOQOERAKVGGQOQINPESNDCKRRAEVGSFIEF
QEKEMEEFVEEREMLIKQEKEMDMKRRHEEILFLEKEFDALBQMKHGLHND
D"
join(25614..25838,26061..26207,26324..26439,26516..26610,
26710..26813,27522..27765,27943..28036,28115..28305,
28385..28604,28688..28855,28944..29092)
/note="gb|AAC97420.1
gene_id:MOML.16
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97243.1"
/db_xref="GI:8809702"
/translation="MAPAGQTAAVAABAAGADAPPQOGRFGSTISGIVRAVFWYF
ASKFEPKQPMPTAPSOQMTPLFKHGSIDMWFTLSBEKENDPNDRLAWHTN
IPYAVWTPESIRPKSLTYVPSFTLNGSLVYAIRPSGSPIDPDREYQPLNSR
THAVATYFPKQKKNKKSILIGSPKDSDESPVEKQDKSDKREYIPVMSIKRN
VTNIVDDFTQFADLVEPTTGYPIYFENEWMLRDKTIPNEVTSBLPLMLEIS
ISMKNQOLFQOVQOSFOMQSYGSMLEDSBDEKFLBENPYLGLITMFSMLHSYF
DPLAFNDILOFNKKNKSMELSAKSVLNLICQFVFLILNDTSMILASGSGVC
IEFWIKGKAMRIEVRSGMIPRLRPHDRSVKATREKYDDIAIKFLSYVLLLVGL
SISLAVERRHSWYSLISLSTSCYVMFGIMPCPOLFITRYKLSVAHLEPROMTYF
LNTIIDLFAVFTKMPILRLHSVFRDVLFLILYQRWYVPVKTRNNEFGSDEET
AEKRLTTEREESBQKTN"
complement(join(29439..29576,29661..29885,29985..30268,
30341..30493,30579..30722,30956..31484))
/note="gene_id:MOML.15
Query Match 100.0%; Score 695; DB 8; Length 81365;
Best Local Similarity 100.0%; Pred. No. 7,8e-131;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CDS
1 GACAAACAAACAAATTAAGCAAGTCATCTGTGAGCAATAATTAATGCGGAAACAA 60
21977 GACAAACAAACAAATTAAGCAAGTCATCTGTGAGCAATAATTAATGCGGAAACAA 22036
61 TTAAGTTAAGCGAAAAAGAAAAAGAGTACAAAAATGAATAACAACTGAAT 120
22037 TTAAGTTAAGCGAAAAAGAAAAAGAGTACAAAAATGAATAACAACTGAAT 22096
121 GAAAAATTTGAGATCGCAATCGAAAAAGAGCGCTTTTGAAGCTTAATAGCTTCTC 180
22097 GAAAAATTTGAGATCGCAATCGAAAAAGAGCGCTTTTGAAGCTTAATAGCTTCTC 22156
181 ATTTGCTCTTCTTGGCATTTATTTCTCTCGAGATCTGACTGACTGACTGACTGACTG 240
22157 ATTTGCTCTTCTTGGCATTTATTTCTCTCGAGATCTGACTGACTGACTGACTGACTG 22216
241 TCTCGGCGCTTTAACTGATTCGCTGCTTGAATCTGTAAGTTTCTGCTTGAAG 300

Db	22217	TCCTCCGCGCTTTAACTTAAGTCTTCCTCCGCTTTACTCGTAAGTTTCTGCTTAGG	22276
QY	301	CCTCCGATCGCTCACCAGCATCTGTGCTCGATTTCTCTTTTCTTCGCTGGAAAA	360
Db	22277	CCTCCGATCGCTCACCAGCATGCAATTCGTGCTCGAATTCCTCTTTTCTTCGCTGGAAAA	22336
QY	361	ATTGCCAATAGTCTCGAATTGCGAAGTTTGTGCTATGAGGTACTTTTTCCTATA	420
Db	22337	ATTGCCCTAATAGTCTCGATTTGCGAAGTTTGTGCTATGAGGTACTTTTTCCTATA	22396
QY	421	TTTTAATGTTCTTAAGTAACGATACCTGCGTTACTGTTTTGTTCATTTGTTGTCT	480
Db	22397	TTTTAATGTTCTTAAGTAACGATACCTGCGCTACTAGTTTTGTTCATTTGTTGTGCT	22456
QY	481	TTCAACGTTTAGTGCTGATGAGGATATTGACCTGTGAAAAAATCCCTGTTTGTGTT	540
Db	22457	TTCAACGTTTAGTGCTGATGAGGATATTGACCTGTGAAAAAATCCCTGTTTGTGTT	22516
QY	541	TTGTTTCATATAAATCGGATGATCTACCTTTGTGCTTGATGTGTTTGTGAGCCT	600
Db	22517	TTGTTTCATATAAATCGGATGATCTACCTTTGTGCTTGATGTGTTTGTGAGCCT	22576
QY	601	ATGCGTTGTGGCTGTATTAATCACTCAAGTGCATGTGGATTTTGAGATTTTGGTAGTG	660
Db	22577	ATGCGTTGTGGCTGTATTAATCACTCAAGTGCATGTGTGGATTTTGAGATTTTGGTAGTG	22636
QY	661	ACTGTGGGTTTCTTTGGTGGCTAATAGGTTGTAATAA	695
Db	22637	ACTGTGGGTTTCTTTGGTGGCTAATAGGTTGTAATAA	22671

RESULT 3			
166494			
LOCUS	166494	7218 bp	DNA
DEFINITION	Sequence 14 from patent US 5670367.		Linear
ACCESSION	166494		
VERSION	166494.1	GI:2724471	
KEYWORDS			
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 7218)		
TITLE	Dorner, F., Scheiflinger, F. and Falkner, F. Gunter.		
JOURNAL	Recombinant fowlpox virus		
FEATURES	Patent: US 5670367-A 14 23-SEP-1997;		
	Location/Qualifiers		
source	1..7218		

ORIGIN

	Query Match	12.3%	Score 85.6	DB 6	Length 7218	
	Best Local Similarity	6.4%	Pred. NO. 1.7e-07			
	Matches 25	Conservative 232	Mismatches 131	Indels 0	Gaps 0	
Qy	174	CTTCCATTTGTCCTTCGTGGTAGTTATTCTTCGCCGAGTCTGACTCACTA	233			
		:::::	::: :::::	::: :::::	::: :::::	
Db	1094	YYY	1153			
Qy	234	CTCTCACTCTCGGCGCTTAAACCTTAGCCTTCCTCGTGTTACTCTGAAGTTTCTGC	293			
		::::: :::::	::: :::::	::: :::::	::: :::::	
Db	1154	YYY	1213			
Qy	294	CTTAGAGCTCCGACGTGCTCACCGAGCATTCGTGTCGATTCTCTTTCTTCGC	353			
		::: :::::	::: :::::	::: :::::	::: :::::	
Db	1214	YYY	1273			
Qy	354	TGAAAAAATGCCCTATGTCCTCGAATTCGAAGTTTTGCTATGGGTACTTTT	413			
		::: :::::	::: :::::	::: :::::	::: :::::	
Db	1274	YYY	1333			
Qy	414	CCCTATAATTTTAGTCTCTTAGGTAAAGATACCTGCGCTTAACTGTTTGTCAATTG	473			
		::::: :::::	::: :::::	::: :::::	::: :::::	

[illegible]

RESULT 4	BT002944	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
	BT002944	2162 bp	Arabidopsis thaliana clone RAFL4-93-K05 (R20243) unknown protein (A55923570) mRNA, complete cds.	BT002944	BT002944.1	GI:27754622	FLI CDNA.	Arabidopsis thaliana (chale cress)	
								Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsids.	
								(Baes I to 2162)	

TITLE
JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 2162)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onders, C.S., Quach, H.L., Tang, C.C., Tortum, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Arabidopsis Full Length cDNA Clones
Unpublished

TITLE
JOURNAL
REFERENCE
AUTHORS

Direct Submission
Submitted (15-JAN-2003) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAPL cDNAs (RAPL cDNA : 'RIKEN
Arabidopsis Full-length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.

The Salk, Stanford, PGEC (SSP) Consortium members carried out the sequencing and annotation of the RFLF cDNAs: Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.T.M., Onders, C.S., Quach, H.L., Tang, C.C., Tortum, M., Wong, C., Wu, H.C., Yu, G., Yuan, S.S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M., Palm, C.J., Shih, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W., Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.

```
FEATURE
  source
    Location/Qualifiers
      1..2162
        /organism="Arabidopsis thaliana"
        /mol_type="mRNA"
        /db_xref="taxon:3702"
        /chromosome="5"
        /clone="RAFL14-93-K05 (R20243)"
        /ecotype="Columbia"
```

```
/note="This clone is in a modified pBluescript vector2
(lambda PS) as a BamHI/XhoI insert."
1. .2162
/gene="At5g23570"
1. .91
/gene="At5g23570"
92. .1969
/gene="At5g23570"
/codon_start=1
/evidence=experimental
/product="unknown protein"
/protein_id="AA022757.1"
/db_xref="GI:27754623"
/translation="MSSRAGPMSEKKNVGGYRPEVQLVQGLAGTLLASODDGGEM
EVISKKKNGKNGTSGKTWVSONSPRAWGGOOGGSGNSVSGKNNVSGKNGNGG
IQANISGRGALSRKYNPNVAPPVSRPFLGEGMNAAGSGHQAIVQFPPEVED
VDNASENDSDALDDSDLDSDYSDVSQSHGSRKONKPKFKFSLDLSLIEQ
INBPQRMHCPCQNGPGLIDWNLHPLAHARTGARVYKHLAEVLEKDLQNGK
ASVIPCGLIYQWKGLEDEKDEYEIVPVMVIMNTRLDKDDNDKMLGNGOELLEYF
DKTEALRARSYSGQGRHSGVLMFESSATGYLEAEHLHELMGLDRJAMGQKRM
FSGGVRLVGLATKODLDFNHSOGKTPKLPKLSYQEMVKEIRQISEDNQOLNY
FRNKLSTQNHAKVYLBESLFINSEKARTADNRIVQRTRMQHEQREMDARDF
MDSIKQIHERDAKENFEMLOQDERAKVVGQOQONINPSSNDCCRRAEVSSFLF
QEKEMEFYVEREMLLKDOEKMEKMDMKRHEBIFLDEKEFDEALBQLMYKHGHMND
D"
1970. .2162
/gene="At5g23570"
/misc_difference 2146
/note="not present in genomic sequence"

ORIGIN
Query Match 12.1%; Score 84; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 4.2e-07;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 199 AGTTATTTCTTCTCCGAGTCTGACTACTCTGCTCCGCGCTTAACT 258
|||||
Db 1 AGTTATTTCTTCTCCGAGTCTGACTACTCTGCTCCGCGCTTAACT 60
|||||

QY 259 TACGTTCTCCGCTTACTCTGT 282
|||||
Db 61 TACGTTCTCCGCTTACTCTGT 84
|||||

RESULT 5
CNS05TCJ
LOCUS 184535 bp DNA linear HTG 06-JUN-2000
DEFINITION Homo sapiens chromosome 14 clone R-691G13, WORKING DRAFT SEQUENCE,
8 ordered pieces.
ACCESSION AL355100
VERSION AL355100.2 GI:9213460
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eureleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 184535)
Genoscope.
Direct Submission
Submitted (06-JUN-2000)
On Jul 15, 2000 this sequence version replaced gi:8346752.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continue. The sequence may be
contaminated with foreign sequence from E.coli, Yeast, vector,
phage, etc. . . even if efforts are made to eliminate these
contaminating sequences. The following BAC sequence is oriented
from the T7 to the Spe end.
Contigs composition :
1386 bp contig from 1 to 1386
37234 bp contig from 1487 to 38720
```

```
1651 bp contig from 38821 to 40471
66410 bp contig from 40572 to 106981
16357 bp contig from 107082 to 123438
2460 bp contig from 123539 to 125998
20897 bp contig from 126099 to 146995
37440 bp contig from 147096 to 184535
-----
Overall quality chart :
Range : bases
0 : 1377
1 - 9 : 1678
10 - 19 : 1569
20 - 29 : 3627
30 - 39 : 13450
40 - 49 : 13488
50 - 59 : 10047
60 - 69 : 22169
70 - 79 : 54644
80 - 89 : 43202
90 - 99 : 19284
-----
Percentage of bases with a quality value >= 40 : 88 %.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 1386: contig of 1386 bp in length
* 1387 1486: gap of 100 bp
* 1487 38720: contig of 37234 bp in length
* 38721 38820: gap of 100 bp
* 38821 40471: contig of 1651 bp in length
* 40472 40571: gap of 100 bp
* 40572 106981: contig of 66410 bp in length
* 106982 107081: gap of 100 bp
* 107082 123438: contig of 16357 bp in length
* 123439 123538: gap of 100 bp
* 123539 125998: contig of 2460 bp in length
* 125999 126098: gap of 100 bp
* 126099 146995: contig of 20897 bp in length
* 146996 147095: gap of 100 bp
* 147096 184535: contig of 37440 bp in length.
*
Location/Qualifiers
1. 184535
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="14"
/clone="R-691G13"
/clone_1b="RPCI-11"

ORIGIN
Query Match 12.0%; Score 83.4; DB 2; Length 184535;
Best Local Similarity 32.3%; Pred. No. 3e-07;
Matches 218; Conservative 0; Mismatches 456; Indels 1; Gaps 1;

QY 2 ACAACAAACAAATTAAGCAAGTCATGTCGTAGCAATTAATTAAGTGGACAAT 61
|||||
Db 38591 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 38650
|||||

QY 62 TAAAGTAAAGCAAAAAAGAAAAAGTACAAAAATGAAAACAAAATCAAACTGAAT 121
|||||
Db 38651 AAAAAAAAAAAAAAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 38710
|||||

QY 122 AAAATTGGAGTCAGAAATCGAAAAACGAGCCGTTTGAAGCTTAATAGCTTCTCA 181
|||||
Db 38711 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 38770
|||||

QY 182 TTGTGCTCTTCTGTCAGTTATTTCTTCTCCGAGTCTGACTACTACTACTACT 241
```

Query Match	11.7%	Score 81.2	DB 6	Length 349980
Best Local Similarity	46.0%	Pred. No. 7.6e-07		
Matches 275	Conservative	0	Mismatches 323	Indels 0
QY	80	AAAAAAAAAGTACAAAATGAAAACAAATCAACATGAATGAAAATTTGGAGTCAGAA	139	
Db	71449	AAAAAAAAAAAAAAAAAAAAAAAAAATTAATATGAAAGTGTTATATTTTAT	71508	
QY	140	TCGAAAAAAGAGCCGTTTGAAGCTTAATAGCTTCCTATTTGCTCTTCGTCA	199	
Db	71509	AAAGTATATTAGAAATAGTTATTCGGTTTTTTTTTTTTTTTTTTTTTTTTTTT	71568	
QY	200	GTTATTTTCTTCCTCCGAGTCGACACTACTCTCAGCTCCGCGCTTAACTT	259	
Db	71569	TTT	71628	
QY	260	ACGTTCCGCTTACTCTGTAAGTTTCTGCTTAGAGCTCCGATCGCCACCGC	319	
Db	71629	TTT	71688	
QY	320	ATGCATTCGTGCTCGATTTCTCTTTCTGCGAGAAAATTCGCCAATGCTCGA	379	
Db	71689	TTT	71748	
QY	380	TTTGAAGGTTTTTGCTCATGAGGTACTTTTTTCCCTAATTTAATGTTCTAGTAA	439	
Db	71749	TTTTTTATT	71808	
QY	440	CGATACCTGCGCTTACTGTTTTTTGTTCAATTTGTGTGCTTCAACCGTTATGCGCTGA	499	
Db	71809	TTTATTTTTTTTTTTT	71868	
QY	500	TGCGAGTATTGACGTGAGAAAACCTCGTTTTTGTGTTTCATTAATAAATCGGA	559	
Db	71869	TTTTTTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	71928	
QY	560	TTCATCTACCTTTTGCTGTGATGTTGTTTTTTGAGCCATAGCTGTGAGCTGTA	619	
Db	71929	TTTTTTTTTTTTTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	71988	
QY	620	TAACTTCAAGTCAAGTGTGATTTTGAGATTTTGTGATGAGTGTGGGTTCTTNG	677	
Db	71989	TTTTTTTTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTGGAGATTTGTTTTGTTAG	72046	
RESULT 7				
AJ592026		1407 bp	DNA	linear
LOCUS				PLN 23-OCT-2003
DEFINITION				Arabidopsis thaliana T-DNA flanking sequence, right border, clone 600D05.
ACCESSION				AJ592026
VERSION				AJ592026.1
KEYWORDS				GI:37941650
SOURCE				right border; T-DNA flanking sequence.
ORGANISM				Arabidopsis thaliana (chale cress)
REFERENCE				Arabidopsis thaliana
AUTHORS				Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.
TITLE				1
JOURNAL				Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechold, N., Crenaud, C., Derose, R., Pelletier, G., Lepiniec, L., Caboche, M. and Lecharny, A.
MEDLINE				T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
PUBMED				EMBO Rep. 3 (12), 1152-1157 (2002)
REFERENCE				2 (bases 1 to 1407)
AUTHORS				Balzerque, S.
TITLE				Direct Submission
JOURNAL				Submitted (23-OCT-2003) Balzerque S., UMRGV, INRA/CNRS, 2 rue

COMMENT

Gaston Cremlieux, 91057 Evry cedex, FRANCE
 PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).

FEATURES

source

```
1. 1407
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Massillawski"
/db_xref="taxon:3702"
/clone="600D05"
/clone_1db="Arabidopsis thaliana T-DNA insertion lines"
misc_feature
1. 1407
/note="T-DNA flanking sequence
right border"
```

ORIGIN

```
Query Match      10.5%; Score 72.8; DB 8; Length 1407;
Best Local Similarity 41.9%; Pred. No. 8.2e-05;
Matches 224; Conservative 0; Mismatches 311; Indels 0; Gaps 0;

QY 157 TTTTGAAGCTTAATAGCTTCCATTTGCTCTCTCGTCACTTATTTCTCTCC 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 703 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 762

QY 217 GGAGTCTGACTACTCTGCTCTCCGCGCTTAACTAGTCTCCGCTTGA 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 763 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 822

QY 277 CTCTGAGTTTCTGCTTGAAGCTTCCGCTCCGCTCAGCAGCATTTCTGCTCGA 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 823 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 882

QY 337 TTTCTCTTTTCTGCTGAAAAATGCCCTAATGTTCTGATTTGAAAGTTTGTG 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 883 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 942

QY 397 CTATGGTACTTTTCCCAATTTTATAGTCTTAGTACGATACCGCTCTTAC 456
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 943 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 1002

QY 457 TGTCTTGTGCTTGTGCTTCAACGTTAGTGGCTGATCGAGATTTGACGT 516
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1003 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 1062

QY 517 GAAAAATCCTGTTTGTGTTTGTTCATATAATGATGATGATGATGATGATG 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1063 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 1122

QY 577 CTATGAGTTTGTGTTTGAAGCTTGTGCTGTGTTAATCACTGATGATGAT 636
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1123 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 1182

QY 637 GTGATTTTGAATTTGTGATGATGATGATGATGATGATGATGATGATGATG 691
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1183 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 1237
```

RESULT 8

AJ591978 1453 bp DNA linear PIN 23-OCT-2003
 LOCUS Arabidopsis thaliana T-DNA flanking sequence, right border, clone
 DEFINITION 598D01.
 ACCESSION AJ591978 GI:37941602
 VERSION

KEYWORDS

SOURCE

right border; T-DNA flanking sequence.
 Arabidopsis thaliana (thale cress)

ORGANISM

Arabidopsis thaliana

REFERENCE

1. Brunaud, V., Balergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Crund, C., Dehose, R., Pelletier, G., Lepoint, L., Caboche, M., and Leclercq, A. T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites EMBO Rep. 3 (12), 1152-1157 (2002)

TITLE

JOURNAL MEDLINE 22363535

AUTHORS

2 (bases 1 to 1453)

REFERENCE

Balergue, S. Direct Submission

TITLE

Submitted (23-OCT-2003) Balergue S., UMRGV, INRA/CNRS, 2 rue

JOURNAL

Gaston Cremlieux, 91057 Evry cedex, FRANCE

COMMENT

PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).

FEATURES

source

```
1. 1453
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Massillawski"
/db_xref="taxon:3702"
/clone="598D01"
/clone_1db="Arabidopsis thaliana T-DNA insertion lines"
misc_feature
1. 1453
/note="T-DNA flanking sequence
right border"
```

ORIGIN

```
Query Match      10.1%; Score 70; DB 8; Length 1453;
Best Local Similarity 43.4%; Pred. No. 0.0003;
Matches 232; Conservative 0; Mismatches 303; Indels 0; Gaps 0;

QY 157 TTTTGAAGCTTAATAGCTTCCATTTGCTCTCTCGTCACTTATTTCTCTCC 216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 390

QY 217 GGAGTCTGACTACTCTGCTCTCCGCGCTTAACTAGTCTCCGCTGTTGA 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 391 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 450

QY 277 CTCTGAGTTTCTGCTTGAAGCTTCCGCTCCGCTCAGCAGCATTTCTGCTCGA 336
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 451 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 510

QY 337 TTTCTCTTTTCTGCTGAAAAATGCCCTAATGTTCTGATTTGAAAGTTTGTG 396
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 511 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 570

QY 571 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 631 TTTTCTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTTNNNTTTT 690

QY 517 GAAAAATCCTGTTTGTGTTTGTTCATATAATGATGATGATGATGATGATG 576
```

[illegible]


```

* 45129 45228: gap of 100 bp
* 46329 46358: contig of 1130 bp in length
* 46359 46458: gap of 100 bp
* 46459 47595: contig of 1137 bp in length
* 47596 47695: gap of 100 bp
* 47696 48785: contig of 1090 bp in length
* 48786 48885: gap of 100 bp
* 48886 49960: contig of 1075 bp in length
* 49961 50060: gap of 100 bp
* 50061 51132: contig of 1072 bp in length
* 51133 51232: gap of 100 bp
* 51233 52356: contig of 1124 bp in length
* 52357 52456: gap of 100 bp
* 52457 53556: contig of 1100 bp in length
* 53557 53656: gap of 100 bp
* 53657 54782: contig of 1126 bp in length
* 54783 54882: gap of 100 bp
* 54883 56008: contig of 1126 bp in length
* 56009 56108: gap of 100 bp
* 56109 57307: contig of 1199 bp in length
* 57308 57407: gap of 100 bp
* 57408 58528: contig of 1122 bp in length
* 58530 58629: gap of 100 bp
* 58630 59799: contig of 1170 bp in length
* 59800 59899: gap of 100 bp
* 59900 60983: contig of 1084 bp in length
* 60984 61083: gap of 100 bp
* 61084 62193: contig of 1110 bp in length
* 62194 62293: gap of 100 bp
* 62294 63397: contig of 1104 bp in length
* 63398 63497: gap of 100 bp
* 63498 64574: contig of 1077 bp in length
* 64575 64674: gap of 100 bp
* 64675 65800: contig of 1126 bp in length
* 65801 65900: gap of 100 bp
* 65901 66993: contig of 1093 bp in length.

```

FEATURES
source 1. 66993
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP13-907D8"
/clone_1b="RP1-13 Human Female BAC"

ORIGIN

Query Match 10.1%; Score 70; DB 2; Length 66993;
Best Local Similarity 41.2%; Pred. No. 0.00018;
Matches 214; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

```

QY 157 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 216
    |||||
DB 56776 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 216
    |||||
QY 217 GGAGCTCTGACCTACTACTCTCCTCCGGCGCTTAACCTACGTTCCCGCTTGA 276
    |||||
DB 56836 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 276
    |||||
QY 277 CTCTGTAAGTTTCTGCTAGAGCTCCGATCGCTCAGCATGATTCGTGCTCGA 336
    |||||
DB 56896 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 336
    |||||
QY 337 TTTCTCTTTTCTGCTGCGTGAATAATGCCCTAATGCTTCGATTCGAAGGTTTGTG 396
    |||||
DB 56956 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 396
    |||||
QY 397 CTATGGTACTTTTCTCCCTATATTTAAGTTCTTAGGTAAGATACCTGCGTCTAC 456
    |||||
DB 57016 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 456
    |||||
QY 457 TGTGTTGCTCATTTGTTGTGCTTCACCGTTAGTGCATGCGAGATATTGACTGT 516
    |||||

```

```

DB 57076 NTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 57135
QY 517 GAAAAATCCTGCTTTTGGCTTTGCTTCATATAAATGGAATGATACCTTTGTG 576
    |||||
DB 57136 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 57195
    |||||
QY 577 CTATGATTTGTTGTTTGAAGCTTATGCTGCTGCTTAACTTCACTGATGT 636
    |||||
DB 57196 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 636
    |||||
QY 637 GTGATTTGGAATTTGTAAGTACTGCTGCTTTCTT 675
    |||||
DB 57256 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 675
    |||||

```

RESULT 10
AJ592058 1434 bp DNA linear PLN 23-OCT-2003
LOCUS Arabidopsis thaliana T-DNA flanking sequence, right border, clone 602D05.
ACCESSION AJ592058.1 GI:37941682
VERSION right border: T-DNA flanking sequence.
KEYWORDS Arabidopsis thaliana (thale cress)
SOURCE Arabidopsis thaliana
ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., Desro, R., Pelletier, G., Lepoint, L., Caboche, M. and Leclercq, A.
T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
EMBO Rep. 3 (12), 1152-1157 (2002)

JOURNAL MEDLINE 22363535
PUBMED 12446565

REFERENCE 2 (bases 1 to 1434)
Balzergue, S.
Direct Submission
Submitted (23-OCT-2003) Balzergue S., UMRGV, INRA/CNRS, 2 rue

COMMENT
Gaston Cremieux, 91057 Evry cedex, FRANCE
PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program "Genoplante" (<http://www.genoplante.com> and <http://genoplante-info.infobiogen.fr>).

FEATURES
source 1. 1434
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Massiliaeskja"
/db_xref="taxon:3702"
/clone="602D05"
/clone_1b="Arabidopsis thaliana T-DNA insertion lines"

misc_feature 1. 1434
/note="T-DNA flanking sequence
right border"

ORIGIN

Query Match 10.0%; Score 69.6; DB 8; Length 1434;
Best Local Similarity 43.7%; Pred. No. 0.00036;
Matches 224; Conservative 0; Mismatches 301; Indels 0; Gaps 0;

```

QY 157 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 216
    |||||
DB 503 TTTTAAAGCTTAATAAGCTTCTCATTTGCTCTCTCTGCTGATTAATTTCTTCCTCC 216
    |||||

```


Db 2225 GTTATTTGTTGTT 2239

RESULT 12
AC022680
LOCUS
DEFINITION
AC022680 88932 bp DNA linear HTG 13-JUL-2000
Homo sapiens chromosome 8 clone RP11-511E14 map 8, LOW-PASS
SEQUENCE SAMPLING.

AC022680
AC022680.2 GI:9143927
HTG: HTGS_PHASE0.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 88932)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 8, clone RP11-511E14
Unpublished

REFERENCE
2 (bases 1 to 88932)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,
Bogunavskiy, L., Bouckhalter, B., Brown, A., Burkett, G., Castle, A.,
Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArrellano, K., Dewar, K., Domino, M., Doyle, M., Fenebor, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, S., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karalas, A., Klein, J.,
Lander, T., Lehoczy, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, K.,
Roy, A., Santos, R., Severny, P., Spencer, B., Stange-Thomann, N.,
Schojnovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6910620.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1998-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence.submissions@genome.wi.mit.edu
Project Information
Center Project name: L5106
Center Clone name: 511_E_14

NOTE: This record contains 86 individual
sequencing reads that have not been assembled into
contigs. Runs of N are used to separate the reads
and the order in which they appear is completely
arbitrary. Low-pass sequence sampling is useful for
identifying clones that may be gene-rich and allows
overlap relationships among clones to be deduced.
However, it should not be assumed that this clone
will be sequenced to completion. In the event that
the record is updated, the accession number will
be preserved.

1 933: contig of 933 bp in length
* 934 1033: gap of 100 bp
* 1034 1961: contig of 928 bp in length
* 1962 2061: gap of 100 bp
* 2062 3005: contig of 944 bp in length
* 3006 3105: gap of 100 bp
* 3106 4066: contig of 961 bp in length
* 4067 4166: gap of 100 bp
* 4167 5120: contig of 954 bp in length

5121 5220: gap of 100 bp
* 5221 6142: contig of 922 bp in length
* 6143 6242: gap of 100 bp
* 6243 7207: contig of 965 bp in length
* 7208 7307: gap of 100 bp
* 7308 8234: contig of 927 bp in length
* 8235 9276: contig of 942 bp in length
* 9277 9377: gap of 100 bp
* 9378 10328: contig of 953 bp in length
* 10329 10429: gap of 100 bp
* 10430 11370: contig of 941 bp in length
* 11371 11470: gap of 100 bp
* 11471 12439: contig of 969 bp in length
* 12440 12539: gap of 100 bp
* 12540 13474: contig of 935 bp in length
* 13475 13574: gap of 100 bp
* 13575 14538: contig of 964 bp in length
* 14539 14638: gap of 100 bp
* 14639 15595: contig of 957 bp in length
* 15596 15695: gap of 100 bp
* 15696 16623: contig of 928 bp in length
* 16624 16723: gap of 100 bp
* 16724 17659: contig of 936 bp in length
* 17660 17759: gap of 100 bp
* 17760 18702: contig of 943 bp in length
* 18703 18802: gap of 100 bp
* 18803 19757: contig of 955 bp in length
* 19758 19857: gap of 100 bp
* 19858 20787: contig of 930 bp in length
* 20788 20887: gap of 100 bp
* 20888 21836: contig of 949 bp in length
* 21837 21935: gap of 100 bp
* 21937 22865: contig of 929 bp in length
* 22866 22965: gap of 100 bp
* 22966 23906: contig of 941 bp in length
* 23907 24006: gap of 100 bp
* 24007 24929: contig of 923 bp in length
* 24930 25029: gap of 100 bp
* 25030 25953: contig of 924 bp in length
* 25954 26054: gap of 100 bp
* 26054 26992: contig of 939 bp in length
* 26993 27092: gap of 100 bp
* 27093 28049: contig of 957 bp in length
* 28050 28149: gap of 100 bp
* 28150 29108: contig of 959 bp in length
* 29109 29208: gap of 100 bp
* 29209 30124: contig of 916 bp in length
* 30125 30224: gap of 100 bp
* 30225 31124: contig of 900 bp in length
* 31125 31224: gap of 100 bp
* 31225 32179: contig of 955 bp in length
* 32180 32279: gap of 100 bp
* 32280 33207: contig of 928 bp in length
* 33208 33307: gap of 100 bp
* 33308 34249: contig of 942 bp in length
* 34250 34349: gap of 100 bp
* 34350 35289: contig of 940 bp in length
* 35290 36327: contig of 938 bp in length
* 36328 36427: gap of 100 bp
* 36428 37353: contig of 926 bp in length
* 37354 37453: gap of 100 bp
* 37454 38408: contig of 955 bp in length
* 38409 38508: gap of 100 bp
* 38509 39427: contig of 919 bp in length
* 39428 39527: gap of 100 bp
* 39528 40459: contig of 932 bp in length
* 40460 40559: gap of 100 bp
* 40560 41483: contig of 924 bp in length
* 41484 41583: gap of 100 bp
* 41584 42543: contig of 960 bp in length
* 42544 42643: gap of 100 bp

42644 43582: contig of 939 bp in length
43583 43682: gap of 100 bp in length
43683 44638: contig of 957 bp in length
44640 44739: gap of 100 bp in length
44740 45670: contig of 931 bp in length
45671 45770: gap of 100 bp in length
45771 46715: contig of 945 bp in length
46716 46815: gap of 100 bp in length
46816 47746: contig of 931 bp in length
47747 47847: gap of 100 bp in length
47847 48772: contig of 926 bp in length
48773 48872: gap of 100 bp in length
48873 49804: contig of 932 bp in length
49805 49904: gap of 100 bp in length
49905 50849: contig of 945 bp in length
50850 50949: gap of 100 bp in length
50950 51842: contig of 893 bp in length
51843 51943: gap of 100 bp in length
51943 52882: contig of 940 bp in length
52883 53928: gap of 100 bp in length
53929 54028: gap of 100 bp in length
54029 54988: contig of 960 bp in length
54989 55088: gap of 100 bp in length
55089 56034: contig of 946 bp in length
56035 56134: gap of 100 bp in length
56135 57063: contig of 925 bp in length
57064 57163: gap of 100 bp in length
57164 58097: contig of 934 bp in length
58098 58197: gap of 100 bp in length
58198 59135: contig of 938 bp in length
59136 59235: gap of 100 bp in length
59236 60161: contig of 926 bp in length
60162 60262: gap of 100 bp in length
60263 61180: contig of 919 bp in length
61181 61280: gap of 100 bp in length
61281 62213: contig of 933 bp in length
62214 63241: gap of 100 bp in length
63242 63341: gap of 100 bp in length
63342 64242: contig of 901 bp in length
64243 64342: gap of 100 bp in length
64343 65301: contig of 959 bp in length
65302 65401: gap of 100 bp in length
65402 66365: contig of 964 bp in length
66366 66465: gap of 100 bp in length
66466 67381: contig of 916 bp in length
67382 67481: gap of 100 bp in length
67482 68455: contig of 974 bp in length
68456 68555: gap of 100 bp in length
68556 69479: contig of 924 bp in length
69480 69579: gap of 100 bp in length
69580 70512: contig of 933 bp in length
70513 71530: gap of 100 bp in length
71531 71630: contig of 918 bp in length
71631 72664: contig of 934 bp in length
72665 73601: gap of 100 bp in length
73601: contig of 937 bp in length

Query Match 9.9%; Score 69; DB 2; Length 88932;
Best Local Similarity 45.0%; Pred. No. 0.00027;
Matches 165; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

325 TTCTGTCGATTCCTCTTTCTCGCGGAAAATGCGCTTAAGTCTCGATTCG 384
68986 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 69045

385 AAGCTTTTTCGATGAGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 444
69046 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 69105

445 CCTCGCTTACTGTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 504

69106 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 69165

505 GTTATTGACGTGAAAAACCTTCGTTTGTGTTTGTTCATATAATCGATTGAT 564

69166 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 69225

565 CTACCTTTGCTGTTGATGTTTGTGTTTGTAGCCTATGCGTTGCTGTATTAAC 624

69226 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 69285

625 TCACGTTACGTCGATTTGATTTGATTTGATGACGTCGCTTCTTTGCTGCTAT 664

69286 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 69345

685 AGGTTGT 691

69346 TTTTCTTTT 69352

RESULT 13
AX346599 6668 bp DNA linear PAT 01-FEB-2002

LOCUS AX346599
DEFINITION Sequence 1670 from Patent WO200928.
ACCESSION AX346599
VERSION AX346599.1 GI:18494485

KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
artificial sequences.

REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Beirlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1670 03-JAN-2002;
EpiGenomics AG (DE)
LOCATION/Qualifiers
source 1. 6668
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically created genomic DNA (Homo sapiens)"

ORIGIN

Query Match 9.9%; Score 68.8; DB 6; Length 6668;
Best Local Similarity 45.8%; Pred. No. 0.00043;
Matches 238; Conservative 0; Mismatches 282; Indels 0; Gaps 0;

156 GTTTTAGAGCTTAATAAGCTTCCTCATTTGCTCTTCTGCGATTAATTTCTTC 215

2742 GTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2801

216 CGAAGTCCTACACACACCTCCTACCTCCGGGCTTTAACTTACCTTCCTGCTT 275

2802 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2861

276 ACCTGTAAGTTTCTGCTTAGAGCTCCGATCGCTACCGCATTCATTCGTGCTG 335

2862 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2921

336 ATTTCCTTTTCTTCGCTGAAAATGCGCTTAAGTCTCGATTTGGAAGTTTGT 395

2222 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2981

396 GCTATGGTACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 455

2982 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3041

456 CTGTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 515

3042 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 3101

516 TGAATAATCTTCGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 575

Db 3102 TGTGTTTGT 3161
 QY 576 GCTTGAATGTTGTTTGTAGCTATGCGTTGTTGCTGTATACCTGACGTCATG 635
 Db 3162 TTTTTCGTT 3221
 QY 636 TGTGATTTTGAATTTGTGAGTGTGAGTGTGCTTTCTTT 675
 Db 3222 TTTTGTGTT 3261

RESULT 14
 AC022851
 LOCUS
 DEFINITION Homo sapiens chromosome 11 clone RP11-284N16 map 11, LOM-PASS
 AC022851 81120 bp DNA linear HTG 13-JUN-2000
 SEQUENCE SAMPLING.
 AC022851
 VERSION AC022851.2 GI:9158172
 KEYWORDS HTG; HTGS PHASE0.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
 AUTHORS Birren, B., Linton, L., Nussbaum, C., Lande, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckert, R., Bedalov, F., Boguslavsky, L., Bouckgalter, B., Brown, A., Burkett, G., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Domino, M., Doyle, M., Fenesor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McSwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 13, 2000 this sequence version replaced gi:6922196.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
 JOURNAL
 COMMENT
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence.submissions@genome.wi.mit.edu
 Project Information
 Center project name: L6170
 Center clone name: 284_N_16

* NOTE: This record contains 82 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1 910: contig of 910 bp in length

* 911 1010: gap of 100 bp
 * 1011 1910: contig of 900 bp in length
 * 1911 2010: gap of 100 bp
 * 2011 2917: contig of 907 bp in length
 * 2918 3017: gap of 100 bp
 * 3018 3921: contig of 904 bp in length
 * 3922 4021: gap of 100 bp
 * 4022 4903: contig of 882 bp in length
 * 4904 5003: gap of 100 bp
 * 5004 5874: contig of 871 bp in length
 * 5875 5974: gap of 100 bp
 * 5975 6860: contig of 886 bp in length
 * 6861 6960: gap of 100 bp
 * 6961 7856: contig of 896 bp in length
 * 7857 7957: gap of 100 bp
 * 7957 8860: contig of 904 bp in length
 * 8861 8960: gap of 100 bp
 * 8961 9822: contig of 862 bp in length
 * 9823 9922: gap of 100 bp
 * 9923 10818: contig of 896 bp in length
 * 10819 10918: gap of 100 bp
 * 10919 11817: contig of 899 bp in length
 * 11818 11917: gap of 100 bp
 * 11918 12797: contig of 880 bp in length
 * 12798 12897: gap of 100 bp
 * 12898 13776: contig of 879 bp in length
 * 13777 13876: gap of 100 bp
 * 13877 14756: contig of 880 bp in length
 * 14757 14856: gap of 100 bp
 * 14857 15788: contig of 882 bp in length
 * 15789 15838: gap of 100 bp
 * 15839 16723: contig of 885 bp in length
 * 16724 16823: gap of 100 bp
 * 16824 17699: contig of 876 bp in length
 * 17700 17799: gap of 100 bp
 * 17800 18699: contig of 900 bp in length
 * 18700 18799: gap of 100 bp
 * 18800 19699: contig of 890 bp in length
 * 19699 19789: gap of 100 bp
 * 19790 20697: contig of 908 bp in length
 * 20698 20797: gap of 100 bp
 * 20798 21638: contig of 841 bp in length
 * 21639 21738: gap of 100 bp
 * 21739 22633: contig of 895 bp in length
 * 22634 22733: gap of 100 bp
 * 22734 23621: contig of 888 bp in length
 * 23622 23721: gap of 100 bp
 * 23721 24617: contig of 896 bp in length
 * 24618 24717: gap of 100 bp
 * 24718 25612: contig of 895 bp in length
 * 25613 25712: gap of 100 bp
 * 25713 26579: contig of 867 bp in length
 * 26580 26579: gap of 100 bp
 * 26580 27577: contig of 898 bp in length
 * 27578 27677: gap of 100 bp
 * 27678 28569: contig of 892 bp in length
 * 28570 28659: gap of 100 bp
 * 28670 29581: contig of 912 bp in length
 * 29582 29681: gap of 100 bp
 * 29682 30578: contig of 897 bp in length
 * 30579 30678: gap of 100 bp
 * 30679 31580: contig of 902 bp in length
 * 31581 31680: gap of 100 bp
 * 31681 32568: contig of 888 bp in length
 * 32569 33562: gap of 100 bp
 * 33563 33562: contig of 894 bp in length
 * 33563 34573: gap of 100 bp
 * 34574 34573: contig of 911 bp in length
 * 34574 35579: gap of 100 bp
 * 35579 35679: contig of 906 bp in length
 * 35680 36547: gap of 100 bp
 * 36547 36547: contig of 868 bp in length

http://genoplante-info.infobiogen.fr.

FEATURES
Location/Qualifiers

1..810
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="Massiliaewskija"
/db_xref="taxon:3702"
/clone="606H03"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
1..810
/note="T-DNA flanking sequence
right border"

ORIGIN

```
Query Match          9.8%; Score 68.4; DB 8; Length 810;
Best Local Similarity 41.2%; Pred.No.0.00069;
Matches 213; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

QY 175 TTCTCATTTGTCCTCTTCCTGTCAGTTATTTCTTCCCGAGTCCGACTCAGTAC 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 142 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTT 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 235 TCTCACTCTCGCGCGCTTAACTAGCTCTCCGCTTACTCTGTAAGTTTCTGCC 294
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 202 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 235 TTAGAGCCCTCCGATCGCCACCGCATGATCTGCTCGATTCTCTTCTCTGCT 354
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 262 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 GGAAAAATTGCCCTAATGTTCTCGATTTCGAAGTTTGTGCTATGSGTTACTTTC 414
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 322 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 381
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 415 CCTATATTTTATAGTCTTAGGTAACGATACCTGCGCTTACTGTTTGTTCATTGT 474
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 382 NTTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 441
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 475 TGTGCTTTCACGCTTAGCGCTGATCGAGATTGACTGTGAAAAATCCTTCGTTT 534
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 442 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 501
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 535 TGGTTTTTGTTCATATAAATCGATGATCTACCTTGTGCTTGAATGTTGTTT 594
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 502 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 561
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 595 GAGCTATGCGTGTGCTGTATTAACCTTCAAGTCAAGTGTGAGATTGAGATT 654
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 562 TTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 621
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 655 GTAGTACTGTGGGTTCTTGTGCTATAGTTGT 691
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 622 TTTTNTTTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNTTNT 658
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: March 29, 2004, 23:22:02
Job time : 2626.78 secs

This Page Blank (uspto)


```

QY 1269 CAAAGAGCCATGATGATGAGAAAGCAATAGTGTTCAGAAAAGTTCTTTGGCAGCTTG 1328
    |||||
Db 2033 CAAAAGAACCCAGAGACTGCTGTAAGAAAGCAATGCTTTAAGATTCTTTTGGAACTTG 1974
QY 1329 GATAGCTTCCATGAGAGATGAAATGAACCAAGAGGAGAGGAGTGTCCAGCTTGT 1388
    |||||
Db 1973 GATGAGCTTATCTATGCAAAAAGATCAATGAAACCGGAAAGGAGTGGACGTGTACAGCTTGT 1914
QY 1389 CAGAAAGAGCTGAGTCCATGATGAGTATTA---ACCTGCAACCTCTAATGCTCATGCG 1445
    |||||
Db 1913 CAGAGTGTGCTCCGCTGCTATGATGTTGTCAGAGAGACTGAGCTCTGATTAATCTCATGCG 1854
QY 1446 AGGACAAAAGAGACTGAGGAGTAAAGCTCATAGAGAAATGGCTGAAGTTTGAAGAG 1505
    |||||
Db 1853 AAAACAAAAGGGGTCAAAAAGGGTGAAGATCCATAGGAGTTTCTGAGCTTTTGGAAAG 1794
QY 1506 GATCTACAGATGAGAGGCGCATCTGTCATCTTGTGTGAGATTAATGGGAGTGGAGAG 1565
    |||||
Db 1793 GAACTGCGAGAAAGGGGCTGTGAGTATTCACCTGGGAGATATTGGTAAAGTGGAAA 1734
QY 1566 GGTTCGGGTGAGAGTGAAGAAAGATTATGAATTTGTGCGCTCCAAATGATCATCATAG 1625
    |||||
Db 1733 GGTTCAAAAGATGAGAGAAAGATCAATGAATAATGTTGGCTCCGATGTTGTCATTCAG 1674
QY 1626 AATACTAGACTGATGAAGAAGCAATACGATTAAGGTGAATTTCTTGTCTTTACTTCTT 1685
    |||||
Db 1673 AATACAAAGCTTGAACAGATGAATAAGTATA----- 1643
QY 1686 TAATTTTCTCTTGATCTTACGATTTTGAATTTGATATGATGAGGCTCGGAGATGG 1745
    |||||
Db 1642 -----AGTGGCTTAGTATGG 1627
QY 1746 CAACCAAGAGCTGTGATATCTTGAAGATGAGGCTTTAGAGACAGCCATTCCTA 1805
    |||||
Db 1626 TAACCAAGAGCTTCTTAATTTATTTAGCGATATGATGCTGTGAAGCTCGACAGCTTA 1567
QY 1806 TGTTCACAGAGGCTATGTGGGATGAGTGTTCGATGTTTGAAGAGAGTCCATGCTGCTA 1865
    |||||
Db 1566 TGGCCCCAGGGTATCGGGGGTGTGAGCGTTTGTATTTTGAAGCATCACTATTAAGTTA 1507
QY 1866 TTGGAGAGCGGAAAGCGCTCCACCGGAGTTAGCTGAGATGGGGTTAGTAAATTTGCTG 1925
    |||||
Db 1506 TCTTGAAGCTGAGCGTCTACACAAAGCACTTTGCAAGAAAGAACTGACGAGATGCTTG 1447
QY 1926 GGGTCAGAGAGCGAGTATGTTTT-----CTGAGAGTGTTCGCCAACTGTATGGCTTCT 1979
    |||||
Db 1446 GTTTAGTCACACCGTATGATTTATTTCTCCGTGTGGCGGAGACAGCTCTATGATTCAT 1387
QY 1980 TGCAAGAGAGAGATCTGACATATTCATCAACACTTTCAGGT 2025
    |||||
Db 1386 GGCATTAAGAAAGAGACCTGACCTTTTCAACAGACATTTGCCAAGT 1341

```

RESULT 2
US-10-424-599-91177

; Sequence 91177, Application US/10424599
; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 91177
; LENGTH: 1377

; TYPE: DNA
; ORGANISM: Glycine max

FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1
US-10-424-599-91177

Query Match 7.7%; Score 253; DB 12; Length 1377;
Best Local Similarity 64.8%; Pred. No. 1.8e-53;
Matches 428; Conservative 0; Mismatches 220; Indels 13; Gaps 3;

```

QY 1047 TATGATTAACACTTTGTGACACCCCACTGATATCTCGCCCTCTTTGGAAGAGATG 1106
    |||||
Db 29 TATGAAGCACTATGATGATCTCAATCTGTAATTTGGCTTCACATGAGATGGCTG 88
QY 1107 AATGGCAGGCAAGAGAGGTTCTGCTCAGCACAGCTGTGAGAGTTTCTGACGTG 1166
    |||||
Db 89 AACTGGCAATCCAGACCTGTGCTGATTCAGTCCATGTAAGGATGAATTTCCACAGAG 148
QY 1167 G-----AGATGATGATGATTAAGCTTCTGAGAAAGAGATGATTCGATGCTTG 1217
    |||||
Db 149 GACCTTCAAAAGATTAATGATGTTGATGATGATGAGGAGAGAGAAATCTGATGATTTG 208
QY 1218 GATGATTCGATGACAGACCTTGCAGATGATTAATGACTCGGATGATGATCAAAAGAG 1277
    |||||
Db 209 GAAGATACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 268
QY 1278 CATGATCAAGAAAGAGATTAAGTGTTCAAAAGTTCTTTGGCAGCTTGTGATGATGCTTG 1337
    |||||
Db 269 CATGAGACAGATGAAGAAAGCAAAATGATTAAGAAATTTCTTGAAGATTTTGAATGGCTTG 328
QY 1338 TCGATGAGCAGATTAATGAATGAACCAAGAGAGAGTGTGCTGACCTTGTCAAGACGA 1397
    |||||
Db 329 ACTGTTGAACAGATTAACAGAACAGAAAGACAGTGGCATTTGTCACAGCTTGCACAGGTGGT 388
QY 1398 CCTGTGCTCATGATTTGATTAAC-----CTGCAACCTCTTACTAGTCAATGAGAGCAAA 1454
    |||||
Db 389 CCTGAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 448
QY 1455 GGAGCTAGGAGAGTAAAGCTCCATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAG 1514
    |||||
Db 449 GATCAAAAAGGATGAAGATTCACAGAGAGCTTGTCTATCTTTGATAGAGAAATTAAGC 508
QY 1515 ATGAGAGGCGCATCTGTCATCTTGTGTGATGATTAATGAGAGAGGTTGGGT 1574
    |||||
Db 509 AAAAGGGGTACTTCAGTAAATTCAGCTGGGGAAGTTTGAAGAGTGAAGGTTTGA 568
QY 1575 GAGATGAAGAAAGATTAATGAATTTCTGCGCTCCCAATGCTATCATCATGATACAG 1634
    |||||
Db 569 GAAAGAG-AAAGAGCAGATGATGATGATTTGGCTCCCAATGCTGATTCAGAAATACAG 627
QY 1635 CTGATTAAGAGAGATTAAGATGATGATGATTTCTTGTCTTTTACTTCTTTAATTTTTC 1694
    |||||
Db 628 CTGGAACAGATGAAGATTAATGATGATGATTAATGATGATTAATTAATCTTGTATATT 687
QY 1695 T 1695
    |||||
Db 688 T 688

```

RESULT 3
US-10-424-599-91176

; Sequence 91176, Application US/10424599
; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 91176

LENGTH: 1589
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
US-10-424-599-91176

Query Match
Best Local Similarity 64.6%; Pred. No. 7e-44; Length 1589;
Matches 358; Conservative 0; Mismatches 184; Indels 12; Gaps 2;

1047 TATGATTAACAAGCTTTGGCAACCCCACTGATCTGCGCTCTTGAAGAAGATGG 1106
29 TATGAAGCACTATGATGACTCAATCTGTAATTCGGCTCCATCGAGCATGGCTGG 88
1107 AATTGGAGGCAAGAGAGGTTCTGCTCAGACACAGCTGTGAGAGTTTCTGACGTG 1166
89 AACTGGCATTCAGACCTGTGTGCAATTCAGTCAATGTAAGGATGAAATTTTCCAGAG 148
1167 G-----AGGATGATGTGATTAATGCTTCTGAGAGAAAGATGATTCGATGCTTTG 1217
149 GACCTTCAAAAGATTAATGATGTTGATGATGATGAGAGAGAGATCTGATGATTTG 208
1218 GATGATTTGATGACGACCTTGAAGTATGATGATGATGATGATGATGATGATGATG 1277
209 GAAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 268
1278 CATGATCAGCAAGCAAGTAAGTGTTCATGAAAGTTCTTTGGCAGCTTGAATGCTGG 1337
269 CATGAGACAGTAAAGAAAGCAATGATTAAGAAAGTTCTTTGAGATTTGATGCTGG 138
1338 TCATCGAGCAGATTAATGAACCAAGAGGAGGATGATGATGATGATGATGATGATGATG 1397
329 ACTGTTGAACAGATTAAGCAAGCAAGAGGATGATGATGATGATGATGATGATGATG 388
1398 CTTGTTGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1454
389 CTGAGAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 448
1455 GAGAGCTAGGAGATTAAGTCTCAGATGATGATGATGATGATGATGATGATGATGATG 1514
449 GAGTCAAAAGAGGATGAAGATTCAGAGGAGCTGATGATGATGATGATGATGATGATG 508
1515 ATGAGAGGAGCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1574
509 AGAAGGAGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 568
1575 GAGGATGAAAGGA 1588
569 GTGACAGCAAGGA 582

RESULT 4
US-10-425-114-18351

Sequence 18351, Application US/10425114
Publication No. US20040034888A1

GENERAL INFORMATION:
APPLICANT: Liu, Jindong

APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.

APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.

APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(5313)B

CURRENT APPLICATION NUMBER: US/10/425,114
NUMBER OF SEQ ID NOS: 73128

SEQ ID NO 18351
LENGTH: 2205
TYPE: DNA

ORGANISM: Zea mays

FEATURE:
OTHER INFORMATION: Clone ID: LIB3079-030_A8_FLI
US-10-425-114-18351

Query Match
Best Local Similarity 55.7%; Pred. No. 2.7e-39; Length 2205;
Matches 518; Conservative 0; Mismatches 330; Indels 82; Gaps 4;

1157 TCCTGACGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1216
452 TTCTGCTCCTGAGGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 511
1217 GATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1276
512 TGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 571
1277 CATGATCAGCAAGCAAGTAAGTGTTCATGAAAGTTCTTTGGCAGCTTGAATGCTT 1336
572 TTTTGAAGCTCGAAGAAAGCAAGGTGTTCAAGAGTTCTTTGAAGTATGATGATGATG 631
1337 GTGATCGAGCAGATTAATGAACCAAGAGGAGGATGATGATGATGATGATGATGATGATG 1396
632 GAGTTGAAACAAATTAATGAACCAAGGATGATGATGATGATGATGATGATGATGATGATG 691
1397 ACCGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1453
692 ACCGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 751
1454 AGAGTGAAGGAGTAAAGTCCATGAGATGATGATGATGATGATGATGATGATGATGATG 1513
752 GGGTTCTCAAGAGGTTAACTTCAAGAGGATGATGATGATGATGATGATGATGATGATGATG 811
1514 GATGAGAGGAGCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1573
812 TCGCAGGAGAACTTCAAGTGTACCAAGGATGATGATGATGATGATGATGATGATGATGATG 869
1574 TGAGATGAAAGATTAATGAATGATGATGATGATGATGATGATGATGATGATGATGATG 1633
870 -CAAGAAAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 928
1634 ACTGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1693
929 CTGGAAGAAAGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 951
1694 CTTTGATCTTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 952
952 -----AGTGAAGGAGCAGTGAAGCAAG 975
1754 AGCTGCTGAATCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1813
976 AGCTCTTGAATTTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1035
1814 AGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1873
1036 CTGGGACACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1095
1874 CCGAAGCTTCAACCGGAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1932
1096 CTGAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1155
1933 --AAGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1990
1156 GCAAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1215
1991 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2050
1216 AGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1275
2051 ATGCTTTAGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2080
1276 GGTCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1305

RESULT 5
US-10-424-599-73594/C
; Sequence 73594, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 73594
; LENGTH: 2890
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_37470C.1
US-10-424-599-73594

Query Match 6.0%; Score 197.8; DB 12; Length 2890;
Best Local Similarity 69.4%; Pred. No. 3.6e-39;
Matches 284; Conservative 0; Mismatches 122; Indels 3; Gaps 1;
QY 1243 GTGATGATTATGACTCGGATGTGATCAAAAGACCTGATTCAGAGACGATTAATGAACAC 1302
DB 2890 GTGATGATTATGACTCGGATGTGATCAAAAGACCTGATTCAGAGACGATTAATGAACAC 2831
QY 1303 GGTTCAAAAAGTCTTTGGGAGCTTGTGATGCTGTGATCGAGCAGATTAATGAACAC 1362
DB 2830 GGTTTAAGGATTTCTTTGAGAACTTGTGATGCTGTGATCGAGCAGATTAATGAACAC 2771
QY 1363 AGAGCAGTGCGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAATGAACAC 1419
DB 2770 AAAGCAGTGCGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAATGAACAC 2711
QY 1420 ACCTGACCTCTACTAGCTCATGCGAGACGAGCAAAAGAGCTAGGCGATTAATGAACAC 1479
DB 2710 GACTGACGCTCTGATTAATGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2651
QY 1480 GAGAAATGCTGATGATTTTGAAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAGAG 1539
DB 2650 GGAAGTTGCTGAGCTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2591
QY 1540 GTGATGATTTATGAGCAGTGAAGAGGTTGGTGAAGATGAAGAAAGATTAATGAACAC 1599
DB 2590 CTGGGGAAGATTTGTAAGTGAAGAGGTTTAAAGAGATGAAGAAAGATTAATGAACAC 2531
QY 1600 TCTGGCCTCCAAATGCTCATGATGATTAATGATTAATGATTAATGATTAATGATTAATG 1648
DB 2530 TTTGGCCTCCAAATGCTCATGATGATTAATGATTAATGATTAATGATTAATGATTAATG 2482

RESULT 6
US-10-424-599-73591/C
; Sequence 73591, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 73591
; LENGTH: 1375

TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1375)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_37468C.1
US-10-424-599-73591

Query Match 5.2%; Score 169.8; DB 12; Length 1375;
Best Local Similarity 69.0%; Pred. No. 3e-32;
Matches 247; Conservative 0; Mismatches 108; Indels 3; Gaps 1;
QY 1243 GTGATGATTATGACTCGGATGTGATCAAAAGACCTGATTCAGAGACGATTAATGAACAC 1302
DB 1375 GTGATGATTATGACTCGGATGTGATCAAAAGACCTGATTCAGAGACGATTAATGAACAC 1316
QY 1303 GGTTCAAAAAGTCTTTGGGAGCTTGTGATGCTGTGATCGAGCAGATTAATGAACAC 1362
DB 1315 GGTTTAAGGATTTCTTTGAGAACTTGTGATGCTGTGATCGAGCAGATTAATGAACAC 1256
QY 1363 AGAGCAGTGCGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAATGAACAC 1419
DB 1255 AAAGCAGTGCGCATTTGTCCAGCTTGTGAGACGACCTGTGATCGATTAATGAACAC 1196
QY 1420 ACCTGACCTCTACTAGCTCATGCGAGACGAGCAAAAGAGCTAGGCGATTAATGAACAC 1479
DB 1195 GACTGACGCTCTGATTAATGATTAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1136
QY 1480 GAGAAATGCTGATGATTTTGAAGAAAGATCTACAGATGAGAGAGAGAGAGAGAGAGAG 1539
DB 1135 GGAAGTTGCTGAGCTTTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1076
QY 1540 GTGATGATTTATGAGCAGTGAAGAGGTTGGTGAAGATGAAGAAAGATTAATGAACAC 1597
DB 1075 CTGGGGAAGATTTGTAAGTGAAGAGGTTTAAAGAGATGAAGAAAGATTAATGAACAC 1018

RESULT 7
US-10-424-599-36215/C
; Sequence 36215, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 36215
; LENGTH: 1103
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1103)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_132704C.1
US-10-424-599-36215

Query Match 3.2%; Score 103.8; DB 12; Length 1103;
Best Local Similarity 67.3%; Pred. No. 1.7e-15;
Matches 175; Conservative 0; Mismatches 83; Indels 2; Gaps 2;
QY 2126 ATTCAATGAGCAAAAGAGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 2184
DB 359 ATTCCAAAGGTAATCTAGATCAAAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 300

```
QY 2185 AAGAGCTGAGCATCTCTGAGACATCAGACGCTGAACCTTTAAGAACAGCTC 2244
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 CACCAATTTGGCAATGATGATGAGACACACAGCTGCTTATTTTGAACAGGCT 240
QY 2245 TCAAAAAGAGACAGACAGCGCAAGTGTGAGAACTCTCTGAAATATGAGGAGAG 2304
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 GTGAAAGTGTCAAAACACACAAAGCTTTGAAGATCTATTTGGTATATGCTGAGAG 180
QY 2305 CTGCGTAGAATCTGACAGGATATCGATGCTGAGACA -GAGAACTAAGATGACAGCTGA 2363
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 CTGCGAAGACAAATGAGAGAAATCGCATTTGTGAGGAGAGAACTAAATGCAACATGA 120
QY 2364 ACAGAACAGGAGAGAGCTAT 2383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 AGAGACCAAAAGNAGATGT 100

RESULT 8
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (379615)
US-10-312-841-2

Query Match      2.5%; Score 81.4; DB 14; Length 3673778;
Best Local Similarity 45.6%; Pred. No. 2.2e-07;
Matches 286; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

QY 80 AAAAAAAAAAGTACAAAATGAAACAAATCAAACTGATGAAAAATTTGAGTCCAGAA 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971449 AAAAAAAAAAAAAAAAAAAAAAGAAAAATTAATGATGAAGAGGTATATATTTTAT 971508
QY 140 TCGAAGAAAGAGGCGCTTTAGAGCTTAATTAAGCTTCTCATTTGTCTCTCGTCA 139
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971509 AAAGTAATTAAGAAATAGTTATCGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971568
QY 200 GTTATATTTCTCTCGGAGTCTGACACTACTCTCACTCTCGGCGCTTTAAACTT 259
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971569 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971628
QY 260 AGCTTCTCGCTCTTACTCTGTAAGTTTCTGCTTAAGAGCTCCGATCCGCTACCGC 319
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971629 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971688
QY 320 ATGATCTGCTGCTCATTTCTTTTCTGCTGGAATAATGCGCTAATGCTCGA 379
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971689 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971748
QY 380 TTTGGAAGTTTTTGCTATGAGGTTACTTTTCCCTAATATTTATAGTTCTTAAGTAA 439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971749 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971808
QY 440 CGATACCTGCGCTTACTGTTTGTTCATTTGTTGTGCTTACCGCTTAAGCTGA 439
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971809 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971868
QY 500 TCGAGATATTTGACTGTGAATAATCCCTGTTTTTGGTTTGTTCATATAATCGA 559
```

```
Db 971869 TTTTATTTTTTTTTTTTTTTAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971928
QY 560 TTGATCTACTTTTGCTTGATGATGTTGTTTGTGAGGCTTAAGCGTTGGCTGTGA 619
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971929 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 971988
QY 620 TAAGTCAAGTCAATGATGATGATTTTGAATTTTGGTAGTCACTGCGGTTCTTGATG 679
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 971989 TTTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 972048
QY 680 GCTATAGTTGTAAAAATGAGTTCTAG 706
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 972049 AGGTTGAAGTGTAGTGTATTATTGTAG 972075

RESULT 9
US-10-424-599-99770/c
; Sequence 99770, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 99770
; LENGTH: 951
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_61108C.1
US-10-424-599-99770

Query Match      2.4%; Score 77; DB 12; Length 951;
Best Local Similarity 63.0%; Pred. No. 1.1e-08;
Matches 119; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 2736 CAGAGCTGAGAGGATGTCATCAAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGGTTGT 2795
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 CAGGTGAGAAATATCTATGTTCTGAAAGCTCAAGACAGAGATGAAGCAATTCGA 547
QY 2796 GGAAGAGAGGAGATGCTGTAATAAGATCAAGAGAAAGATGAGACATGAAGAGAG 2855
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 GCGAGAGAGAGAGAAATTAATAAATTCATGAAGAGAAAGTTGGCTCTGAAGAGAA 487
QY 2856 GCATCAAGAGAGATATTTGATCTGAGAGAAAGATTTGATGAGCTTTGGAACGCTCAT 2915
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 486 GCAGTGGCAAGAGCAGGTTGAACCTGAGAAAGAGATTAGAGAAATGAAGTACGACGCTCAT 427
QY 2916 GTACAGCA 2924
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 426 GGACAAAGTA 418

RESULT 10
US-10-311-455-1670
; Sequence 1670, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter-
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
```

```

? DE 10013058.8
? DE 10019173.8
? DE 10032529.7
? DE 10043826.1
? PRIOR FILING DATE: 2001-04-06
? 2000-04-06
? 2000-04-07
? 2000-06-30
? 2000-09-01
? NUMBER OF SEQ ID NOS: 228
? SEQ ID NO 52
? LENGTH: 9539
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: chemically created genomic DNA (Homo sapiens)
US-10-239-676-52

```

Query Match	2.0%	Score 66.8;	DB 14;	Length 9539;
Best Local Similarity	46.7%	Pred. No. 2e+05;		
Matches 212;	Conservative	0;	Mismatches 242;	Indels 0;
				Gaps 0

QY	251	TTTAACTTAAGCTTC	CCGTCGTTACT	CTGTAAGTTTCTG	CTAAGAGCCCGATCG	310
Db	202	TTTTTTTTTTTGTAT	TTTTTGTATTA	TTTTTGAGTTTTTTTTTT	TTTATTTTTTTTTTTC	261
QY	311	CCGCAACCGATCAT	TTGTGCTCGATTC	CTCTTTTTCTTCG	CTGGAATAATGCCCTTA	370
Db	262	GTTTTTTTTTTTT	GTGCTTTGTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTATTT	321
QY	371	TGTTCTGATTCGA	AGGTTTTTTTG	GCATACGGTACTTTTT	CCCAATATTTAGTT	430
Db	322	TTATTTGTTTTTAT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTT	381
QY	431	CTAAGTAACGAT	CACTGGCTTACTG	TTTTTTTGTCA	TTTGTGTGCTTACCGTT	490
Db	382	TTTTTTTGTTTTTTT	TGTTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTT	441
QY	491	AGTCGCTATCGGA	GTATTGACTGTGA	AAATCCCTGTTTTTG	TTTTTGTGTTTCA	550
Db	442	ATTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTATTTT	501
QY	551	TAAATCGAATGAT	CTACCTTTTGCT	TGATGATGTTTTTT	TGAGCCATAGCGTTT	610
Db	502	TTTTTTTTTTTTTT	TGATTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTT	561
QY	611	GGCTGTATTA	ACTTCAAGTCA	TGTCATGTCGATTT	TGAACTTTGGTAGCTGCGTT	670
Db	562	TTTTTTTTTTTGT	TTATTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTTTTTTTTTTTTTTTTTTTTTTT	TTTTTTATTTATTT	621
QY	671	TCTTTGGGCTAT	AGGTGTGA	AAAAATGAGTTCT	704	
Db	622	TTATTTATATAT	TTTTTTTTTTTGAG	TAGGATG	655	

```

RESULT 12
US-10-240-453-54
; Sequence 54, Application US/10240453
; Publication No. US20030148326A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIRENBERG, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
; TITLE OF INVENTION: Transcription
; TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associate
; TITLE OF INVENTION: with DNA Transcription
; FILE REFERENCE: 5013.1009
; CURRENT APPLICATION NUMBER: US/10/240.453
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03973
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
;

```

```

RESULT 13
US-10-424-599-57885
: Sequence 57885, Application US/10424599
: Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
: TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
: SEQ ID NO 57885
LENGTH: 673
:
TYPE: DNA
ORGANISM: Glycine max
FEATURE:

```

```

RESULT 14
US-10-311-455-1931
Sequence 1931, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
APPLICANT: BERLIN, Kurt
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining the Methylation of Cytosine
TITLE OF INVENTION: Cytosine methylation
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1931
LENGTH: 14006
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
FEATURE:
NAME/KEY: unsure
LOCATION: 8289, 8310, 8313
OTHER INFORMATION: n is a or g or c or t
US-10-311-455-1931
Query Match 2.0%; Score 66; DB 14; Length 14006;

```

Best Local Similarity 45.3%; Pred. No. 4.2e-05; Indels 0; Gaps 0;
Matches 240; Conservative 0; Mismatches 290;

```

QY 157 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 216
    |||
Db 1974 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2033
    |||
QY 217 GGAATCCTGACCTACCTACCTCCTCGGCGCTTAACTAGCTCTCCGTCGTTA 276
    |||
Db 2034 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2093
    |||
QY 277 CTCTGAGTTTCTGCTTACGAGCCGCTCAGCCGATTCATCTGCTCGA 336
    |||
Db 2094 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2153
    |||
QY 337 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 396
    |||
Db 2154 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2213
    |||
QY 397 CTATGAGTTTCTGCTTACGAGCCGCTCAGCCGATTCATCTGCTCGA 456
    |||
Db 2214 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2273
    |||
QY 457 TGTCTTGTGCTTGTGCTTACGCTTACGCTTACGCTTACGCTTACGCTTAC 516
    |||
Db 2274 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2333
    |||
QY 517 GAAAAATCCTGCTTGTGCTTACGCTTACGCTTACGCTTACGCTTACGCTTAC 576
    |||
Db 2334 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2393
    |||
QY 577 CTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 636
    |||
Db 2394 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2453
    |||
QY 637 GTGATTTTGAATTTGTGCTTACGCTTACGCTTACGCTTACGCTTACGCTTAC 686
    |||
Db 2454 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 2503
    |||

```

RESULT 15
US-10-311-455-2147
; Sequence 2147, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPERBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; TITLE OF INVENTION: Cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2147
; LENGTH: 113515
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; US-10-311-455-2147

Query Match 2.0%; Score 65.8; DB 14; Length 113515;
Best Local Similarity 48.5%; Pred. No. 0.0002; Indels 0; Gaps 0;
Matches 181; Conservative 0; Mismatches 192;

QY 325 TTCTGTGCTGATTTCTTTCTTCCTGCTGAAAAATGCTTAAGTTCTGATTTGCG 384

```

Db 51453 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 51512
    |||
QY 385 AAGTTTGTGCTAAGGTTACTTTTCCCTAATTTTAAGTTCTAGTAAGAGA 444
    |||
Db 51513 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 51572
    |||
QY 445 CTGCGCTTACGTTTGTGCTTACGCTTACGCTTACGCTTACGCTTACGCTTAC 504
    |||
Db 51573 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 51632
    |||
QY 505 GTATTTGACCTGAAAAATCCTGCTTGTGCTTACGCTTACGCTTACGCTTACGCTTAC 564
    |||
Db 51633 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 51692
    |||
QY 565 CTACCTTTGTGCTTACGTTTGTGCTTACGCTTACGCTTACGCTTACGCTTACGCTTAC 624
    |||
Db 51693 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 51752
    |||
QY 625 TCAGCTTACGTTGAGATTTGAGATTTGAGATTTGAGATTTGAGATTTGAGATTTGAGATTT 684
    |||
Db 51753 TTTTAAAGCTAATAGAGCTCCATTTGCTCTTCGTCAGTTATTTCTTCCTCC 51812
    |||
QY 685 AGTTGTAATAAT 697
    |||
Db 51813 AAGTTTATGAGAT 51825
    |||

```

Search completed: March 30, 2004, 08:57:32
Job time : 1121.16 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2004, 16:31:11 ; Search time 246.409 Seconds
(without alignments)
7375.802 Million cell updates/sec

Title: US-10-030-829-1

Perfect score: 3275
Sequence: 1 gcaacaacaacaataa.....caaatataggtacaat 3275

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PTCUR.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.6	2.6	7218	1 US-08-232-463-14	Sequence 14, Appl
2	74.4	2.3	272	4 US-09-313-294A-121	Sequence 121, App
3	72.2	2.2	274	4 US-09-313-284A-463	Sequence 463, App
4	68.6	2.1	7218	1 US-08-232-463-14	Sequence 14, Appl
5	52.6	1.6	7286	3 US-09-331-581-3	Sequence 3, Appl
6	52.6	1.6	7938	3 US-09-331-581-14	Sequence 14, Appl
7	50.6	1.5	396	4 US-09-640-173-53	Sequence 53, Appl
8	50.6	1.5	396	4 US-09-733-550-53	Sequence 53, Appl
9	50.6	1.5	10619	4 US-10-204-708-4	Sequence 4, Appl
10	50.2	1.5	5562	4 US-10-204-708-63	Sequence 63, Appl
11	49.4	1.5	19124	2 US-08-487-826B-13	Sequence 13, Appl
12	49.2	1.5	6040	4 US-10-204-708-69	Sequence 69, Appl
13	48.6	1.5	51259	3 US-08-781-891-209	Sequence 209, App
14	48.6	1.5	51259	4 US-09-618-166-209	Sequence 209, App
15	47.8	1.5	2447	2 US-09-014-969-14	Sequence 14, Appl
16	47.4	1.4	5666	4 US-10-204-708-29	Sequence 29, Appl
17	47.4	1.4	8607	4 US-10-204-708-72	Sequence 72, Appl
18	47.4	1.4	19233	4 US-10-204-708-45	Sequence 45, Appl
19	47.2	1.4	8961	4 US-10-204-708-79	Sequence 79, Appl
20	47	1.4	6617	4 US-09-976-594-268	Sequence 268, App
21	46	1.4	3489	2 US-08-728-323A-1	Sequence 1, Appl
22	46	1.4	3489	4 US-09-238-568-1	Sequence 1, Appl
23	46	1.4	3489	4 US-09-410-359-1	Sequence 1, Appl
24	46	1.4	32207	2 US-08-770-379-20	Sequence 20, Appl
25	46	1.4	32207	3 US-08-757-669A-20	Sequence 20, Appl
26	46	1.4	32207	4 US-09-230-371A-20	Sequence 20, Appl
27	45.6	1.4	289	3 US-09-007-005-17	Sequence 17, Appl

C 28	45.6	1.4	289	3 US-09-244-796-17	Sequence 17, Appl
C 29	45.4	1.4	740	3 US-08-998-416-563	Sequence 563, App
C 30	45.4	1.4	11049	4 US-10-204-708-23	Sequence 23, Appl
C 31	44.4	1.4	9347	4 US-10-204-708-36	Sequence 36, Appl
C 32	44.4	1.4	16442	3 US-08-781-891-208	Sequence 208, App
C 33	44.4	1.4	16442	4 US-09-618-166-208	Sequence 24, Appl
C 34	44.2	1.3	1493	1 US-08-340-820-24	Sequence 24, Appl
C 35	44.2	1.3	1493	1 US-08-593-535-24	Sequence 24, Appl
C 36	44	1.3	240	1 US-08-628-417-6	Sequence 6, Appl
C 37	44	1.3	2394	4 US-09-800-729-33	Sequence 33, Appl
C 38	44	1.3	11050	4 US-10-204-708-85	Sequence 85, Appl
C 39	44	1.3	19233	4 US-10-204-708-46	Sequence 46, Appl
C 40	43.6	1.3	359	4 US-09-621-976-16008	Sequence 16008, A
C 41	43.6	1.3	2915	4 US-09-336-115C-5	Sequence 5, Appl
C 42	43.6	1.3	5501	4 US-10-204-708-38	Sequence 38, Appl
C 43	43.6	1.3	11015	4 US-10-204-708-55	Sequence 55, Appl
C 44	43.4	1.3	140	1 US-08-628-417-5	Sequence 5, Appl
C 45	43.4	1.3	359	4 US-09-621-976-16019	Sequence 16019, A

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESS: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pC-Fls
US-08-232-463-14
Query Match 2.6%; Score 85.6; DB 1; Length 7218;

Best Local Similarity 6.4%; Pred. No. 1,6e-12;
Matches 25; Conservative 232; Mismatches 131; Indels 0; Gaps 0;

```
QY 174 CTTCCTCATTTGCTCTCTCTCTGTCAGTTAATTTCTCTCCGAGTCTGACTCACTA 233
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1094 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1153

QY 234 CTCCTACTCCGGGCGTTTAAGTTACGTTCCGCTGTTACTCTGTAAGTTTCTGCG 293
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1154 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1213

QY 294 CTTAGAGCTCCGATCCGCTCCGAGTATCTGCTGCTGCTCTCTTTCTCTGCG 353
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1214 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1273

QY 354 TGGAAAAATGCGCTTAATGTTCTGAGTTGCAAGGTTTTGCTGTAAGGTTACTTTT 413
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1274 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1333

QY 414 CCCATATTTTATAGTCTTATAGTATGAGATACCGCTGCTACTGTTTGTCTATTTG 473
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1334 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1393

QY 474 TTGCTTTTACCGCTTATGCTGATCGAGATTTGACTGAAAAATCTCTGTTT 533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1394 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1453

QY 534 TTGCTTTTGTCTTATTAATCGGATT 561
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1454 CTTTACTACTTGATGATGATGTTAATT 1481
```

RESULT 2

```
US-09-313-294A-121
; Sequence 121, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laljudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 121
; LENGTH: 272
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700548569H1
; NAME/KEY: unsure
; LOCATION: 14, 32, 61, 127, 162, 192
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-121
```

Query Match 2.3%; Score 74.4; DB 4; Length 272;
Best Local Similarity 58.7%; Pred. No. 2.4e-10;
Matches 142; Conservative 0; Mismatches 97; Indels 3; Gaps 1;

```
QY 1777 TATGAGGCTCTTAAGACGCGCATTTCTATGTCACAGGGGCGATCGGGAGTATGTT 1836
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TATGAGCAAGTNAAGACGTCATGCTTATGTCCTCTGCGCACCGGTGATATAGGCTN 61

QY 1837 CTGATGTTTGAAGACAGTGCATGCTATTTGAGAGCGCAACGCTCAACCGGAGTTA 1896
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 TTATATTTTGAAGACTGAGTGGGCTACATGAGAGTGAAGTCACTGCAATAAACACTTT 121

QY 1897 GCTGAGATGGGGTATGATGATTTGCTCG--GATCAAGACCGAGTATGTTTCTGGA 1953
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 GTTANCAAGTACAGACAGGAATTCATGCACTAGCAGCAGTTCATTTGTGCTGCT 181
```

```
QY 1954 GGTGTTGCCACTGATGCTTCTTCTGCAACGAGACAGATCTGACATATTCAATCA 2013
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 GGGAAAAAGGNCCTATATGTTTCTTATGCAAAAGAGAGATATGACGATTTAAACAG 241

QY 2014 CA 2015
    |||||
Db 242 CA 243
```

RESULT 3

```
US-09-313-294A-463
; Sequence 463, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
; APPLICANT: Laljudi, Raghunath V.
; APPLICANT: Ito, Laura Y.
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
; FILE REFERENCE: PL-0017 US
; CURRENT APPLICATION NUMBER: US/09/313,294A
; CURRENT FILING DATE: 1999-05-14
; NUMBER OF SEQ ID NOS: 7600
; SOFTWARE: PERL Program
; SEQ ID NO 463
; LENGTH: 274
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. 6476212 700549278H1
; NAME/KEY: unsure
; LOCATION: 15, 53
; OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-463
```

Query Match 2.2%; Score 72.2; DB 4; Length 274;
Best Local Similarity 60.6%; Pred. No. 9.6e-10;
Matches 152; Conservative 0; Mismatches 95; Indels 4; Gaps 2;

```
QY 1777 TATGAGGCTCTTAAGACGCGCATTTCTATGTT-CCACAGGGGCGATGTTGGATGAGT 1835
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TATGAGCAAGTNAAGACGCTATGCTTATGTCCTCTGCGGACCGGTGATATGACGCT 61

QY 1836 TCTGATCTTGAAGACAGTCCACTGCTATTTGAGAGCCGAGCCTTCCACCGGAGTT 1895
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 62 GTTATATTTTGAAGGCTCCAGCTGTGGCTACATGGAAGCTGACGTCTGATTAACACTT 121

QY 1896 AGCTGAGATGGGGTATGATGAATTTGCTG--GGTCAGAAAGGACAGTATGTTTCTGG 1952
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 122 TGTTATCAAGGTACAGACAGGAATTCATGCACTTACGACCTTACGAGGTTGATTTGCTGG 181

QY 1953 AGGTGTTGCCAACTGATGCTTCTTGTGCAAGAGCAAGATCTGACATATTCAATCA 2012
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 182 TGGGAAAAAGCAACTATATGTTTCTTACGAAAACAAAGAGATATGAGGCAATTTAACAA 241

QY 2013 ACACTCTCAAG 2023
    |||||
Db 242 GCATTGCCAG 252
```

RESULT 4

```
US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHRIFFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
```

STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pc-Fls
US-08-232-463-14

Query Match 2.1%; Score 68.6; DB 1; Length 7218;
Best Local Similarity 9.0%; Pred. No. 6,1e-08;
Matches 41; Conservative 230; Mismatches 184; Indels 0; Gaps 0;

2731 TTTCGACAGCTGAGAGAGTGTCAAGCTTCATCGAGTTTCAAGAGAGAGATGAGAG 2790
1442 TTTCGTCRRR 1363
2791 TTTCGACAGAGAGAGAGTGTCAAGATTAAGATCAAGAGAGAGATGAGAGATGAG 2850
1382 RR 1323
2851 AAGAGCATGACGAGAGATATTGATCTGAGAGAAAGATTGATGAGGCTTGGACAG 2910
1322 RR 1263
2911 CTCATGACAGATGAGCTTCACATGAGATGATGAGACAAAGCTGTGACAGAG 2970
1262 RR 1203
2971 ACAAGACTAAGTTCTTTGTTGCTTTGGTATGCGAAAAGTAGAGATCTGAGAGAC 3030
1202 RR 1143
3031 TCATTAATAATACAGACAATACTAAGAGATTATAGATTATATCTCCAAATTTAG 3090
1142 RR 1063
3091 TAGACGATCTAAGAGAGATTAGTTCTGTGACTAAACCAAGTTCTTAGATTTT 3150
1082 RR 1023
3151 GTTTTGTGTAATAATTCTATATGAAAGTTAGA 3185
1022 GTGAGCTATGCGAAGAGAGAAATAAGTTATA 988

RESULT 5
US-09-331-581-3
Sequence 3, Application US/09331581
Patent No. 6130070
GENERAL INFORMATION:
APPLICANT: TOHDA, Hideki
APPLICANT: HAMA, Yuko
APPLICANT: KUMAGAI, Hiromichi
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
FILE REFERENCE: 0059-1142-0PCT
CURRENT APPLICATION NUMBER: US/09/331,581
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/JP98/04929
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: JP 9-314608
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 3
LENGTH: 7286
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-3

Query Match 1.6%; Score 52.6; DB 3; Length 7286;
Best Local Similarity 56.7%; Pred. No. 0.0013;
Matches 97; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

9 AACCAAAATTAGCAAGTCTGTCGTAATTAATTAATAGTGGAGCAATTAAGTTA 68
4559 AACATACATGAGAGAAATGCTATGAGCTGAGAAATTAATGACGTCACAGCAAAA 4618
69 AGCGAAAGAGAAAAAAGGTACAAAATGAAAACAAATCAACTGAATGAATTT 128
4619 AA 4678
129 GAGCTCCAGATCGGAAAAAGAGCGCGTTTGAAGCTTAATTAAGCTTCT 179
4679 TACTTTGAGCGCGGAAAGAACGCGGATCCAGACATGATTAAGATCAT 4729

RESULT 6
US-09-331-581-14
Sequence 14, Application US/09331581
Patent No. 6130070
GENERAL INFORMATION:
APPLICANT: TOHDA, Hideki
APPLICANT: HAMA, Yuko
APPLICANT: KUMAGAI, Hiromichi
TITLE OF INVENTION: INDUCTION PROMOTER GENE AND SECRETORY SIGNAL GENE
TITLE OF INVENTION: USABLE IN SCHIZOSACCHAROMYCES POMBE, EXPRESSION VECTORS
FILE REFERENCE: 0059-1142-0PCT
CURRENT APPLICATION NUMBER: US/09/331,581
CURRENT FILING DATE: 1999-06-30
EARLIER APPLICATION NUMBER: PCT/JP98/04929
EARLIER FILING DATE: 1998-10-30
EARLIER APPLICATION NUMBER: JP 9-314608
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 14
LENGTH: 7938
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:DNA
US-09-331-581-14

Query Match 1.5%; Score 52.6; DB 3; Length 7938;
Best Local Similarity 56.7%; Pred. No. 0.0013;
Matches 97; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 9 AACAAATTAAGCAAGTCAATGTTCTGTCGCAATTAATTAATTAAGTGAACAATTAAGTGA 68
DB 5211 AACATATCATGAGAAAGATGCTATGTAGTGAATAATTAATGAATGACACAGACAAAA 5270
QY 69 AGCGAAAAAGAAAAAGATGTCACAAAATGAAAACAAATCAACTGAATGAAAAATTT 128
DB 5271 AAAG 5330
QY 129 GAGTCGAGATCGAAAAACGAGCGCTTTAGAGCTTAATAGCTTCC 179
DB 5331 TACCTCTGAGCGGAAAAAGAACGACCGGATCCAGACATGATAGATACAT 5381

RESULT 7

US-09-640-173-53
Sequence 53, Application US/09640173
Patent No. 6613515
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Stoik, John A.
TITLE OF INVENTION: OVARIAN TUMOR SEQUENCES AND
FILE REFERENCE: 210121.484C2
CURRENT APPLICATION NUMBER: US/09/640.173
CURRENT FILING DATE: 2000-08-15
NUMBER OF SEQ ID NOS: 196
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 53
LENGTH: 396
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(396)
OTHER INFORMATION: n = A,T,C or G
US-09-640-173-53

Query Match 1.5%; Score 50.6; DB 4; Length 396;
Best Local Similarity 49.8%; Pred. No. 0.00079;
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 448
DB 18 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 77
QY 449 CGTCTTACTGTTTGTTCATTTTGTGTGCTTACCGTTAGTCGTCATCGAGATAT 508
DB 78 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 137
QY 509 TTGACGTGAAAAATCCTGCTTTTGGTTTGTTCATATAAATGAGATGATAC 568
DB 138 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 197
QY 569 CTTTGTGCTTGAATGTTTGTGAGCCATGCGTGTGCTGTATTAACCTCA 627
DB 198 TTTTGTGCTTGAATGTTTGTGAGCCATGCGTGTGCTGTATTAACCTCA 256

RESULT 8

US-09-713-550-53
Sequence 53, Application US/09713550
Patent No. 6617109
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Stoik, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.484C4

CURRENT APPLICATION NUMBER: US/09/713,550
CURRENT FILING DATE: 2000-11-14
NUMBER OF SEQ ID NOS: 205
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 53
LENGTH: 396
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(396)
OTHER INFORMATION: n = A,T,C or G
US-09-713-550-53

Query Match 1.5%; Score 50.6; DB 4; Length 396;
Best Local Similarity 49.8%; Pred. No. 0.00079;
Matches 119; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 389 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 448
DB 18 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 77
QY 449 CGTCTTACTGTTTGTTCATTTTGTGTGCTTACCGTTAGTCGTCATCGAGATAT 508
DB 78 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 137
QY 509 TTGACGTGAAAAATCCTGCTTTTGGTTTGTTCATATAAATGAGATGATAC 568
DB 138 TTTTGTGCTATGAGTACTTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTG 197
QY 569 CTTTGTGCTTGAATGTTTGTGAGCCATGCGTGTGCTGTATTAACCTCA 627
DB 198 TTTTGTGCTTGAATGTTTGTGAGCCATGCGTGTGCTGTATTAACCTCA 256

RESULT 9

US-10-204-708-4
Sequence 4, Application US/10204708
Patent No. 6677731
GENERAL INFORMATION:
APPLICANT: OLEK, Alexander
APPLICANT: PIEPENBROCK, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA Replication
TITLE OF INVENTION: by Assessing DNA Methylation
FILE REFERENCE: 5013.1012
CURRENT APPLICATION NUMBER: US/10/204.708
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: PCT/EP01/03971
PRIOR FILING DATE: 2001-04-06
PRIOR APPLICATION NUMBER: DE 10019058.8
PRIOR FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: DE 10019173.8
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 4
LENGTH: 10619
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-4

Query Match 1.5%; Score 50.6; DB 4; Length 10619;
Best Local Similarity 49.8%; Pred. No. 0.0055;
Matches 128; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 404 TTAATTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTGCTTACTGTTT 463
DB 128 TTAATTTTCCCTATATTTATAGTCTTAGGTAAAGATACCTGCTTACTGTTT 463

Db 15736 TTAATTTAATAATTTTTTCTTTTTTTTTTTGTTTATGACATATAATTTTTTTTTTTT 15677

Oy 617 TTAACTTCACGTCATGCTGGATTGAGATTTGGTACGACGTGGGTTTC 675

Db 15676 AATGTTTTTTTTCTCTCTTGGTTTAAATTTTTTTTAAATCAATTTTTTTTTTAAT 15618

RESULT 12
US-10-204

```

US-10-204-708-69
; Sequence 69, Application US/10204708
; Patent No. 6677731
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis Associated with DNA Replication
; TITLE OF INVENTION: by Assessing DNA Methylation
; FILE REFERENCE: 5013.1012
; CURRENT APPLICATION NUMBER: US/10/204,708
; CURRENT FILING DATE: 2003-05-06
; PRIOR APPLICATION NUMBER: PCT/EP01/03971
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 98
; SEQ ID NO 69
; LENGTH: 6040
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-204-708-69

```

Query Match	1.5%	Score 49.2	DB 4	Length 6040
Similarity	49.5%	Pred. No. 0.0094		
Best Local				
Matches 186; Conservative	0	Mismatches 163	Indels 7	Gaps 2

US-08-781-891-209/c

: Sequence 209, Application US/08781891
 : Patent No. 6090620
 :
 : GENERAL INFORMATION:
 : APPLICANT: Fu, Ying-Hui
 : APPLICANT: Yu, Chang-Sh
 : APPLICANT: Oshima, Junko
 : APPLICANT: Mulligan, John T.
 : APPLICANT: Schellenberg, Gerald D.
 : TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
 : TITLE OF INVENTION: WARNER'S SYNDROME
 : NUMBER OF SEQUENCES: 209
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: SEED and BERRY LLP
 : STREET: 6300 Columbia Avenue, 701 Fifth Avenue
 : CITY: Seattle
 : STATE: Washington
 : COUNTRY: USA
 : ZIP: 98104-7092
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : OPERATING SYSTEM: IBM PC compatible
 : SOFTWARE: Patentin Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/781,891
 : FILING DATE: 27-DEC-1996
 : CLASSIFICATION: 800
 : ATTORNEY/AGENT INFORMATION:
 : NAME: No. 6090620tenburg Ph. D., Carol
 : REGISTRATION NUMBER: 39,317
 : REFERENCE/DOCKET NUMBER: 240052.419
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (206) 622-4900
 : TELEFAX: (206) 682-6031
 : INFORMATION FOR SEQ ID NO: 209:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 51259 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear

Query Match	1.5%;	Score 48.6;	DB 3;	Length 51259;
Similarity	52.2%;	Pred. 0.048		
Best Local	Matches 108;	Conservative	0;	Mismatches 99;
			Indels 0;	Gaps 0;

RESULT 14

US-09-618-166-209/C
Sequence 209, Application US/09618166
Patent No. 6583112
GENERAL INFORMATION:
APPLICANT: Fu, Ying-Hui
Yu, Chang-En
Oshima, Junko
Mulligan, John T.

```

1 Schellenberg, Gerald D.
2 TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
3 WERNER'S SYNDROME
4
5 NUMBER OF SEQUENCES: 209
6
7 CORRESPONDENCE ADDRESS:
8 ADDRESSEE: Seed Intellectual Property Law Group
9 STREET: 701 Fifth Avenue, Suite 6300
10 CITY: Seattle
11 STATE: Washington
12 COUNTRY: USA
13 ZIP: 98104-7092
14
15 COMPUTER READABLE FORM:
16 MEDIUM TYPE: Floppy disk
17 COMPUTER: IBM PC compatible
18 OPERATING SYSTEM: PC-DOS/MS-DOS
19 SOFTWARE: PatentIn Release #1.0, Version #1.30
20
21 CURRENT APPLICATION DATA:
22 APPLICATION NUMBER: US/09/618,166
23 FILING DATE: 17-Jul-2000
24 CLASSIFICATION: <Unknown>
25
26 ATTORNEY/AGENT INFORMATION:
27 NAME: Mcmasters, David D.
28 REGISTRATION NUMBER: 33,963
29 REFERENCE/DOCKET NUMBER: 240052.419C1
30
31 TELECOMMUNICATION INFORMATION:
32 TELEPHONE: (206) 622-4900
33 TELEFAX: (206) 682-6031
34
35 INFORMATION FOR SEQ ID NO: 209:
36
37 SEQUENCE CHARACTERISTICS:
38 LENGTH: 51259 base pairs
39 TYPE: nucleic acid
40 STRANDEDNESS: single
41 TOPOLOGY: linear
42
43 SEQUENCE DESCRIPTION: SEQ ID NO: 209:
44
45 US-09-618-166-209
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967

```

1 CORRESPONDENCE ADDRESS:
2 ADDRESSEE: Genetics Institute, Inc.
3 STREET: 87 CambridgePark Drive
4 CITY: Cambridge
5 STATE: MA
6 COUNTRY: U.S.A.
7 ZIP: 02140
8
9 COMPUTER READABLE FORM:
10 MEDIUM TYPE: Floppy disk
11 COMPUTER: IBM PC compatible
12 OPERATING SYSTEM: PC-DOS/MS-DOS
13 SOFTWARE: Patentn Release #1.0, Version #1.30
14 CURRENT APPLICATION DATA:
15 APPLICATION NUMBER: US/09/014,969
16 FILING DATE:
17 CLASSIFICATION:
18 ATTORNEY/AGENT INFORMATION:
19 NAME: Sprunger, Suzanne A.
20 REGISTRATION NUMBER: 41,323
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: (617) 498-8284
23 TELEFAX: (617) 876-5851
24 INFORMATION FOR SEQ ID NO: 14:
25 SEQUENCE CHARACTERISTICS:
26 LENGTH: 2447 base pairs
27 TYPE: nucleic acid
28 STRANDEDNESS: double
29 TOPOLOGY: linear
30 MOLECULE TYPE: cDNA
31 US-09-014-969-14

Query Match	1.5%;	Score 47.8;	DB 2;	length 2447;
Best Local Similarity	47.2%;	Pred. No. 0.013;		
Matches 142;	Conservative 1;	Mismatches 158;	Indels 0;	Gaps 0;
QY	390	TTTGTGCTATGGGTTACTTTTTTCCCTATATTTTATAGTCTTAAAGTAAAGATACCTGC	449	
Db	2447	TT	2388	
QY	450	GCTCTACTGTTTGTTCATTTTGTGTGCTTTACCGTTTAGTCGCTGATCGAGTAT	509	
Db	2387	TT	2328	
QY	510	TGACTGTGAAAAATCCTTCGTTTTTGGTTTTTGGTTCATATAATCGATTTGATCTAC	569	
Db	2327	TT	2268	
QY	570	TTTGTGCTTGTAGTGTGTTTTTGAAGCCATAGCGTTGTGCGCTGTATATACCTACG	629	
Db	2267	TTTTTTTTTTTTTTTTTTTTTTTTTTTCCCAAAAGTTTATTGTACATTTAAAGTACAAA	2208	
QY	630	TTCAATGTGAGATTTTAGATTTTGGATGTGACTGTGGGTTTCTTTGGTGCCTATAGTT	689	
Db	2207	ATCAATAACACAGATCCAGATATGTGAACAATATATACATATCATACAGCCATTATTTA	2148	
QY	690	G 690		
Db	2147	G 2147		

Search completed: March 30, 2004, 03:14:55
Job time : 249.409 secs

This Page Blank (uspto)

XX Beclin C, Elmayer T, Vaucheret H;
XX
XX WPI; 2001-159529/16.
XX
XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
PT resistance in plants and, when inhibited, for increasing transgene
PT expression.
XX
XX Claim 1; Page 31-32; 36pp; French.
XX
XX The present sequence represents the genomic sequence of the Arabidopsis
CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
CC inactivation (degradation of RNA) and for resistance to viruses.
CC Overexpression of SGS3 results in plants with increased resistance to
CC viruses, while inactivation of SGS3 in transgenic plants (e.g. by
CC expressing antisense RNA, by mutation or by homologous recombination)
CC increases the level of the transgene product. This product may e.g.
CC impart resistance to herbicide, insects or pathogens), alter contents of
CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an
CC immunoglobulin or interferon
XX
XX Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;
S0
Query Match 100.0%; Score 3275; DB 4; Length 3275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GACAAACAAACAAATTTAGCAATGTCATGTTCTGACATTAATTAATAGTGGAAACA 60
DB 1 GACAAACAAACAAATTTAGCAATGTCATGTTCTGACATTAATTAATAGTGGAAACA 60
QY 61 TTAAGTTAAGGAG 120
DB 61 TTAAGTTAAGGAG 120
QY 121 GAAATTTGAGAGTCAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
DB 121 GAAATTTGAGAGTCAGAGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
QY 181 ATTTGCTCTTCTTCTGTCAGTTATTTTCTCTCCGAGTCTGACTCACTACTCTC 240
DB 181 ATTTGCTCTTCTTCTGTCAGTTATTTTCTCTCCGAGTCTGACTCACTACTCTC 240
QY 241 TCTCGGAGCTTTAACTTAAGTTCTCGGAGTTACTCTGTAAGTTTCTGCTGAGAG 300
DB 241 TCTCGGAGCTTTAACTTAAGTTCTCGGAGTTACTCTGTAAGTTTCTGCTGAGAG 300
QY 241 TCTCGGAGCTTTAACTTAAGTTCTCGGAGTTACTCTGTAAGTTTCTGCTGAGAG 300
DB 241 TCTCGGAGCTTTAACTTAAGTTCTCGGAGTTACTCTGTAAGTTTCTGCTGAGAG 300
QY 301 CCTCGATCGCTCAACGATGATCTGCTGATTTCTTTTCTGCTGAGAGAGAGAGAGAG 360
DB 301 CCTCGATCGCTCAACGATGATCTGCTGATTTCTTTTCTGCTGAGAGAGAGAGAGAG 360
QY 361 ATTGCGCTTAATGTTCTGATTTGAGAGTTTGTGCTAGAGAGTTTCTTTTCTGCTA 420
DB 361 ATTGCGCTTAATGTTCTGATTTGAGAGTTTGTGCTAGAGAGTTTCTTTTCTGCTA 420
QY 421 TTTTATAGTCTTGAAGTACGATCTGCTGATTTGCTGATTTGCTGATTTGCTGAT 480
DB 421 TTTTATAGTCTTGAAGTACGATCTGCTGATTTGCTGATTTGCTGATTTGCTGAT 480
QY 481 TTCAACGTTTATGCTGATGAGATTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAG 540
DB 481 TTCAACGTTTATGCTGATGAGATTTGATGATGAGAGAGAGAGAGAGAGAGAGAGAG 540
QY 541 TTGTTTCAATTAATCGAGATGATCTTTTGTGCTTGAATGTTTGTGAGAGCT 600
DB 541 TTGTTTCAATTAATCGAGATGATCTTTTGTGCTTGAATGTTTGTGAGAGCT 600
QY 601 ATGGGTTGAGCTTTGATTAATCTCAAGTCAATGAGAGAGAGAGAGAGAGAGAGAGAG 660
DB 601 ATGGGTTGAGCTTTGATTAATCTCAAGTCAATGAGAGAGAGAGAGAGAGAGAGAGAG 660
QY 661 ACTGTGGGTTCTTTGTGAGCTATAGTTGTAATAAGAGTTCTAGAGAGCTGCTCAATGT 720

DB 661 ACTGTGGGTTCTTTGTGAGCTATAGTTGTAATAAGAGTTCTAGAGAGCTGCTCAATGT 720
QY 721 CTAAAG 780
DB 721 CTAAAG 780
QY 781 TGGAG 840
DB 781 TGGAG 840
QY 841 AGAAG 900
DB 841 AGAAG 900
QY 901 ATCTCTTAAGAGCTTTGAG 960
DB 901 ATCTCTTAAGAGCTTTGAG 960
QY 961 GAAACAAATGATTCGAG 1020
DB 961 GAAACAAATGATTCGAG 1020
QY 1021 GTGAG 1080
DB 1021 GTGAG 1080
QY 1081 CTGAGCTCTCTTTGAG 1140
DB 1081 CTGAGCTCTCTTTGAG 1140
QY 1141 CAGCTGTGAG 1200
DB 1141 CAGCTGTGAG 1200
QY 1201 ATGATTCGATGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
DB 1201 ATGATTCGATGCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
QY 1261 ATGAG 1320
DB 1261 ATGAG 1320
QY 1321 GCAAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
DB 1321 GCAAGCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
QY 1381 CAGCTTGTGAG 1440
DB 1381 CAGCTTGTGAG 1440
QY 1441 ATGAG 1500
DB 1441 ATGAG 1500
QY 1501 AAAAGAGATCTACAGATGAG 1560
DB 1501 AAAAGAGATCTACAGATGAG 1560
QY 1561 GGAAGGTTTGGTGAAGATGAAAGAGATTAAGATTTGTGCTGCTCAATGATGATCA 1620
DB 1561 GGAAGGTTTGGTGAAGATGAAAGAGATTAAGATTTGTGCTGCTCAATGATGATCA 1620
QY 1621 TCATGAATTAATCAAGATGAG 1680
DB 1621 TCATGAATTAATCAAGATGAG 1680
QY 1681 TTCTTTAATTTTCTCTGATTTCTAATGATTTAGAGATTAATGATTTAGAGAGAGAGAGAG 1740
DB 1681 TTCTTTAATTTTCTCTGATTTCTAATGATTTAGAGATTAATGATTTAGAGAGAGAGAGAG 1740
QY 1741 ATGGGCAACCAAGAGCTGCTGAATCTTGACAGATGAGAGCTTTAGAGAGAGAGAGAGAGAG 1800

Db 1741 ATGGGCAACCAAGAGCTGCTGGAAATCTTCACAAAGTATGAGGCTCTTAGAGCAGCCAT 1800
 Qy 1801 TCCATAGGTCCAGAGGGGCTCATGCGGATGAGTGTTCGATGTTTGAGAGAGGCCACT 1860
 Db 1801 TCCATAGGTCCAGAGGGGCTCATGCGGATGAGTGTTCGATGTTTGAGAGAGGCCACT 1860
 Qy 1861 GGGTATTTGGAGGCGGAGCGGCTCCAGCGGAGTTAGCTGAGATGGGTTAGATGAAT 1920
 Db 1861 GGGTATTTGGAGGCGGAGCGGCTCCAGCGGAGTTAGCTGAGATGGGTTAGATGAAT 1920
 Qy 1921 GCTTGGGGTCAGAAAGCGCATGATGTTTCTGAGAGTGTTCGCAACTGTATGCTTCTT 1980
 Db 1921 GCTTGGGGTCAGAAAGCGCATGATGTTTCTGAGAGTGTTCGCAACTGTATGCTTCTT 1980
 Qy 1981 GCAAGCAAGCAAGATCTGAGATATTCATCAACCTCAAGTCTCTCCCAAGGA 2040
 Db 1981 GCAAGCAAGCAAGATCTGAGATATTCATCAACCTCAAGTCTCTCCCAAGGA 2040
 Qy 2041 AATTGATATATGCTTTAGTTTGTCTATGGAATTTAAAGTTTGTGTCGGTAA 2100
 Db 2041 AATTGATATATGCTTTAGTTTGTCTATGGAATTTAAAGTTTGTGTCGGTAA 2100
 Qy 2101 TGCACTGTTATGATATATCTATGATTCATTAGGCAAAACAAGCTGAAATTCAGATTG 2160
 Db 2101 TGCACTGTTATGATATATCTATGATTCATTAGGCAAAACAAGCTGAAATTCAGATTG 2160
 Qy 2161 AAATCATACCAAGAGATGTTGTAAGGAGCTGAGGCGATCTCTGAGGACATCAGAG 2220
 Db 2161 AAATCATACCAAGAGATGTTGTAAGGAGCTGAGGCGATCTCTGAGGACATCAGAG 2220
 Qy 2221 CTGAACACTTTTAAAGAACAGCTCTCAAAAAGAAAGCAAGCAGGAGGCTTGAGGA 2280
 Db 2221 CTGAACACTTTTAAAGAACAGCTCTCAAAAAGAAAGCAAGCAGGAGGCTTGAGGA 2280
 Qy 2281 TCTCTGGAATTTAGAGGAGAGAGCTGCTGTAAGAACTGCAAGATATCGATCTGAGA 2340
 Db 2281 TCTCTGGAATTTAGAGGAGAGAGCTGCTGTAAGAACTGCAAGATATCGATCTGAGA 2340
 Qy 2341 CAGAGACTAAGATGAGCAGATGAGCAAGAAAGGAGATATTTCTTGAAAT 2400
 Db 2341 CAGAGACTAAGATGAGCAGATGAGCAAGAAAGGAGATATTTCTTGAAAT 2400
 Qy 2401 CACAACTTGACATTTTGTATTTACCTACTGATTCATATTTGATTAATGTCACAA 2460
 Db 2401 CACAACTTGACATTTTGTATTTACCTACTGATTCATATTTGATTAATGTCACAA 2460
 Qy 2461 AAAAAGCTGCTGTTGTAAGATGATGACACAGCAGGTTTTCATGATTCATCAAA 2520
 Db 2461 AAAAAGCTGCTGTTGTAAGATGATGACACAGCAGGTTTTCATGATTCATCAAA 2520
 Qy 2521 CAGATCATGAAAGAGAGCGCAAGAGAGAAATTCGAGATGTTGCGACACAGGA 2580
 Db 2521 CAGATCATGAAAGAGAGCGCAAGAGAGAAATTCGAGATGTTGCGACACAGGA 2580
 Qy 2581 CGTGCCAAAGTGTGTCAGCAGCAGCAGCAACATTAATCCCTCTAGCAATGAGATTGC 2640
 Db 2581 CGTGCCAAAGTGTGTCAGCAGCAGCAGCAACATTAATCCCTCTAGCAATGAGATTGC 2640
 Qy 2641 CGAAAGAGTATATGATCTAACTAACTAACTAACTAACTAACTAACTAACTAACTAA 2700
 Db 2641 CGAAAGAGTATATGATCTAACTAACTAACTAACTAACTAACTAACTAACTAACTAA 2700
 Qy 2701 AGAGTACTGAATTAATCCGCTTTGATTTCTTTCGAGAGCTGAGAGAGTCAAGCTTC 2760
 Db 2701 AGAGTACTGAATTAATCCGCTTTGATTTCTTTCGAGAGCTGAGAGAGTCAAGCTTC 2760
 Qy 2761 ATCGAGTTTCAAGAGAGAGATGAGAGATTTTGAGAGAGAGAGAGATGATTAATA 2820
 Db 2761 ATCGAGTTTCAAGAGAGAGATGAGAGATTTTGAGAGAGAGAGAGATGATTAATA 2820
 Qy 2821 GATCAAGAGAGAGAGATGAGAGATCAAGAGAGAGAGAGATCAAGAGATATTTGATCTG 2880
 Db 2821 GATCAAGAGAGAGAGATGAGAGATCAAGAGAGAGAGAGATCAAGAGATATTTGATCTG 2880

Qy 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCAATGTAACAGCATGAGGCTTCACATGAA 2940
 Db 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCAATGTAACAGCATGAGGCTTCACATGAA 2940
 Qy 2941 GATGATTTGAGCAAAAGTCTGCTGATACAAAGCAAGACTAATGTTCTTTGTTGGCTTTG 3000
 Db 2941 GATGATTTGAGCAAAAGTCTGCTGATACAAAGCAAGACTAATGTTCTTTGTTGGCTTTG 3000
 Qy 3001 GTATGTCGGAAGAGAGATCTGAGAGCTCCATTTAATATCTAGCAAAATCTAAGGA 3060
 Db 3001 GTATGTCGGAAGAGAGATCTGAGAGCTCCATTTAATATCTAGCAAAATCTAAGGA 3060
 Qy 3061 GATTATGATTAATATCTCCATTTTATGAGAGCGGATCTAAGGAAGCATTAAGTCTT 3120
 Db 3061 GATTATGATTAATATCTCCATTTTATGAGAGCGGATCTAAGGAAGCATTAAGTCTT 3120
 Qy 3121 GTGACTAAACCAAGTTTCTTATGATTTGTTTGTGTAATAATTCATATGAAAG 3180
 Db 3121 GTGACTAAACCAAGTTTCTTATGATTTGTTTGTGTAATAATTCATATGAAAG 3180
 Qy 3181 TTAGACATTTTACCAAAAGCTGAGGATGATACAGATGCAATCAAAATCATGTTTTT 3240
 Db 3181 TTAGACATTTTACCAAAAGCTGAGGATGATACAGATGCAATCAAAATCATGTTTTT 3240
 Qy 3241 AGAATTTTATATCTCAAAAATATATGGGTCAAT 3275
 Db 3241 AGAATTTTATATCTCAAAAATATATGGGTCAAT 3275

RESULT 2
 ID AAF25374 standard; cDNA; 1878 BP.
 XX AAF25374;
 AC 15-MAY-2001 (first entry)
 DT 15-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of the Arabidopsis SGS3 polypeptide.
 XX
 KW SGS3 gene; post-transcriptional inactivation; RNA degradation;
 KW viral resistance; resistance; fatty acid content; protein content; ss.
 OS Arabidopsis thaliana.
 XX
 FH Key
 FT CDS
 FT 1.1878
 FT /*cag= a
 FT /product= "SGS3"
 XX
 EN W0200105951-A2.
 XX
 XX 25-JAN-2001.
 XX
 PF 13-JUL-2000; 2000MO-FR002052.
 XX
 PR 16-JUL-1999; 99FR-00009417.
 PR 26-JAN-2000; 2000FR-00001006.
 XX
 PA (AVET) AVENTIS CROPS SCIENCE SA.
 PA (INRG) INST NAT RECH AGRONOMIQUE.
 XX
 PI Beclin C, Elmayan T, Vaucheret H;
 XX
 DR WPI; 2001-159529/16.
 DR P-PSDB; AAB31798.
 XX
 PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
 PT resistance in plants and, when inhibited, for increasing transgene
 PT expression.
 XX
 PS Claim 1; Page 32-35; 36pp; French.
 XX

CC The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.
 CC The SGS3 gene is essential for post-transcriptional inactivation
 CC (degradation of RNA) and for resistance to viruses. Overexpression of
 CC SGS3 results in plants with increased resistance to viruses, while
 CC inactivation of SGS3 in transgenic plants (e.g. by expressing antisense
 CC RNA, by mutation or by homologous recombination) increases the level of
 CC the transgene product. This product may e.g. impart resistance (to
 CC herbicide, insects or pathogens), alter contents of essential fatty acids
 CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or
 CC interferon

XX Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 44.6%; Score 1462; DB 4; Length 1878;
 Best Local Similarity 83.3%; Pred. No. 0;

Matches 1878; Conservative 0; Mismatches 0; Indels 376; Gaps 4;

QY 696 ATGAGTTCTAGGGCTGCTCAATGTCTAAAGAAAGAAAGCTTCAGGGTGTATAGGCTT 755
 Db 1 ATGAGTTCTAGGGCTGCTCAATGTCTAAAGAAAGAAAGCTTCAGGGTGTATAGGCTT 60
 QY 756 GAGGTGAACAGTGGTTCAGAGTTTGGAGAGGACAGACTGGCTTTTCAAGATGAT 815
 Db 61 GAGGTGAACAGTGGTTCAGAGTTTGGAGAGGACAGACTGGCTTTTCAAGATGAT 120
 QY 816 GAGAGAGTGGAGGCTATTTCAAGAGAAACAAGAACAGAAACAGAAACCTTCTGGA 875
 Db 121 GAGAGAGTGGAGGCTATTTCAAGAGAAACAAGAACAGAAACAGAAACCTTCTGGA 180
 QY 876 AAAAAGTGGGTTTCTCAAGATTCGAATCTCTAGAGCTTGGGGTGTCTAGAGCAAGG 935
 Db 181 AAAAAGTGGGTTTCTCAAGATTCGAATCTCTAGAGCTTGGGGTGTCTAGAGCAAGG 240
 QY 936 AAGAGTAGCAAGTATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATG 995
 Db 241 AAGAGTAGCAAGTATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGCAATG 300
 QY 996 CCGGGCAATCAAGCTAACATATCTGTGGGGACAGAGCTTGAAGAAATGATGATAC 1055
 Db 301 CCGGGCAATCAAGCTAACATATCTGTGGGGACAGAGCTTGAAGAAATGATGATAC 360
 QY 1056 AACCTTGGGCAACCCCACTGATCTGGCTCTCTTGAAGAGAGATGAAATTGGCAG 1115
 Db 361 AACCTTGGGCAACCCCACTGATCTGGCTCTCTTGAAGAGAGATGAAATTGGCAG 420
 QY 1116 GCAGAGAGAGTCTGTCTAGACACACAGCTGTGAGAGTTCCTGCACTGAGAGATGAT 1175
 Db 421 GCAGAGAGAGTCTGTCTAGACACACAGCTGTGAGAGTTCCTGCACTGAGAGATGAT 480
 QY 1176 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 1235
 Db 481 GTGATTAATGCTTCTGAGAGAGAGATGATTCGATGCTTTGGATGATTTCTGATGACGAC 540
 QY 1236 CTTGCAAGTGAATTTATGATCGAGATGAGTCAAAAAGCCATGATGATGATGATGATGAT 1295
 Db 541 CTTGCAAGTGAATTTATGATCGAGATGAGTCAAAAAGCCATGATGATGATGATGATGAT 600
 QY 1296 AATAAGTGGTCAAAAAGTCTTTGGAGAGCTTGGATGATGATGATGATGATGATGAT 1355
 Db 601 AATAAGTGGTCAAAAAGTCTTTGGAGAGCTTGGATGATGATGATGATGATGATGAT 660
 QY 1356 GAACACAGAGAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1415
 Db 661 GAACACAGAGAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
 QY 1416 TATAAAGTGAACCTCTACTAGCTCAGTCAAGAGCAAAAAGAGCTTAGGCGATTAAGCTC 1475
 Db 721 TATAAAGTGAACCTCTACTAGCTCAGTCAAGAGCAAAAAGAGCTTAGGCGATTAAGCTC 780
 QY 1476 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATTTAAGATGAGAGGCGATCTGTCAAT 1535
 Db 781 CATAGAGAAATGGCTGAAGTTTGAAGAAAGATTTAAGATGAGAGGCGATCTGTCAAT 840

QY 1536 CTTGTGTGAGATTTATGAGCACTGAGAGGTTTGGGTGAGAGATGAAAGATTTATGAA 1595
 Db 841 CTTGTGTGAGATTTATGAGCACTGAGAGGTTTGGGTGAGAGATGAAAGATTTATGAA 900
 QY 1596 ATTGTGCGCTCCAAATGCTATCATATGAAATCTGATGATGATGATGATGATGATGAT 1655
 Db 901 ATTGTGCGCTCCAAATGCTATCATATGAAATCTGATGATGATGATGATGATGATGAT 960
 QY 1656 AAGGTGAATCTTCTGTCTTTACTTCTTAATTTTCTTGAATTCATGATCTTA 1715
 Db 961 A----- 961
 QY 1716 GAATGTTAATTTGATGCTCGCATGGCAACAAGAGCTCTGAAATCTTGACAA 1775
 Db 962 -----AGTGGCTCGCATGGCAACAAGAGCTCTGAAATCTTGACAA 1007
 QY 1776 GTATGAGGCTCTTAGAGACAGCAATTCCTATGCTTCAAGGGGCACTGTGGATGATGCT 1835
 Db 1008 GTATGAGGCTCTTAGAGACAGCAATTCCTATGCTTCAAGGGGCACTGTGGATGATGCT 1067
 QY 1836 TCTGATGTTTGAAGACAGTGCACCTGGCTATTTTGAAGGCGCAAGCCTCCACCGGAGTT 1895
 Db 1068 TCTGATGTTTGAAGACAGTGCACCTGGCTATTTTGAAGGCGCAAGCCTCCACCGGAGTT 1127
 QY 1896 AGCTGAGATGGGTTGATGAAATTCCTGGGCTCAAGACCGCATGATGTTTCTGAGG 1955
 Db 1128 AGCTGAGATGGGTTGATGAAATTCCTGGGCTCAAGACCGCATGATGTTTCTGAGG 1187
 QY 1956 TGTTCGCAACTGATGAGCTTCTTCAAGAGCAAGATCTGACATTTCAATCAACA 2015
 Db 1188 TGTTCGCAACTGATGAGCTTCTTCAAGAGCAAGATCTGACATTTCAATCAACA 1247
 QY 2016 CTCTCAAGGTTCTCTCCCAAGAAATTTGATATGCTTTAGTTTGTCAATGGAAT 2075
 Db 1248 CTCTCA----- 1253
 QY 2076 TTAAGTTTGTTCGCTCGGTTAATGATCTGTTATGATATATCATGATTCATTAGG 2135
 Db 1254 -----AAG 1256
 QY 2136 CAAGACAGGCTGAATTTGAGTTGAATCATACAGAGATGTTGTAAGAGCTGAG 2195
 Db 1257 CAAGACAGGCTGAATTTGAGTTGAATCATACAGAGATGTTGTAAGAGCTGAG 1316
 QY 2196 GCAATCTCTGAGAGCAATCAGCAGCTGAATCTTTAAGAACAGCTCTCAAAACAGAA 2255
 Db 1317 GCAATCTCTGAGAGCAATCAGCAGCTGAATCTTTAAGAACAGCTCTCAAAACAGAA 1376
 QY 2256 CAAGACAGGCTGAATTTGAGTTGAATCATACAGAGATGTTGTAAGAGCTGAG 2315
 Db 1377 CAAGACAGGCTGAATTTGAGTTGAATCATACAGAGATGTTGTAAGAGCTGAG 1436
 QY 2316 TGCAGAGATTAATCGATCTGAGACAGAGAACTTAAGATGACATGAACAGAGG 2375
 Db 1437 TGCAGAGATTAATCGATCTGAGACAGAGAACTTAAGATGACATGAACAGAGG 1496
 QY 2376 AAGAGTATGATTTTCTAGAAAATCAGAACTTGATTTGATTTACTATGATTTCA 2435
 Db 1497 AG----- 1498
 QY 2436 CATTTTGAATTAATGTTCAACAAAAAAGCTGTGTGTTGAAGATGATGACAG 2495
 Db 1499 -----AGATGATGACAG 1514
 QY 2496 CAGTTTTTCAATGATTTCAATCAACAGATTCATGAAGAGAGAGAGAGAGAG 2555
 Db 1515 CAGTTTTTCAATGATTTCAATCAACAGATTCATGAAGAGAGAGAGAGAGAG 1574
 QY 2556 TTTGAGATGTTGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2615
 Db 1575 TTTGAGATGTTGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1674
 QY 2616 TAACTCTCTAGCAATGACATGCTCCGAAAGAGGTATATGATTAATCAATATCCT 2675

```

Db      ||||| 1635 TAAATCCCTCTAGCATGACGATGCGCAAG----- 1665
Qy      2676 CTGGCGTTTTTGTTCCTAACCTAAGAGTAACTGATTTCCGGTTTGATTTCTTCG 2735
Db      1666 ----- 1665
Qy      2736 CAGAGCTGAGAGAGTGTCAAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGAGTTGT 2795
Db      1666 -AGAGCTGAGAGAGTGTCAAGCTTCATCGAGTTTCAAGAGAAAGATGAGAGAGTTGT 1724
Qy      2796 GGAAGAGAGAGAGTGTCTGATTAAGATCAAGAGAGAGATGAGAGAGATGAGAGAG 2855
Db      1725 GGAAGAGAGAGAGTGTCTGATTAAGATCAAGAGAGAGATGAGAGAGATGAGAGAGAG 1784
Qy      2856 GCATCAGAGAGAGATTTTGTATCTGAGAGAAAGATTTGATGAGCTTTGAGAGAGCTCAT 2915
Db      1785 GCATCAGAGAGAGATTTTGTATCTGAGAGAAAGATTTGATGAGCTTTGAGAGAGCTCAT 1844
Qy      2916 GTACAGAGATGCGCTTCACATGAGATGATTGA 2949
Db      1845 GTACAGAGATGCGCTTCACATGAGATGATTGA 1878

RESULT 3
ABX81661
ID      ABX81661 standard; cDNA; 272 BP.
XX      AC      ABX81661;
XX      DT      24-APR-2003 (first entry)
XX      DE      Corn ear-derived polynucleotide (cpd) #121.
XX      KW      Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;
XX      KW      structural gene; functional gene; regulatory gene;
XX      KW      corn ear-specific profile; gene transcription; gene expression;
XX      KW      hybrid plant; desirable trait expression; plant breeding program;
XX      KW      inheritance; desirable characteristic; growth; development; yield;
XX      KW      disease resistance; environmental adaptability; quality; yield;
XX      KW      multigene trait; plant; gene; ss.
XX      OS      Zea mays.
XX      PN      US6476212-B1.
XX      PD      05-NOV-2002.
XX      PF      14-MAY-1999; 99US-00313294.
XX      PR      26-MAY-1998; 98US-0086722P.
XX      PA      (INCY-) INCYTE GENOMICS INC.
XX      PI      Lalagudi RV, Ito LY, Sherman BK;
XX      DR      WPI; 2003-208840/20.
XX      PT      Novel purified corn-ear derived polynucleotide useful as hybridization
XX      PT      probe for detecting polynucleotide in sample, and for identifying,
XX      PT      evaluating, and altering desired characteristics associated with growth,
XX      PT      development.
XX      PS      Example; SEQ ID NO 121; 390bp; English.
XX      CC      The present invention relates to the isolation of corn ear-derived
XX      CC      polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022
XX      CC      and SATMON023. Some of the cdp's uniquely identify structural, functional,
XX      CC      and regulatory genes of corn ear. The polynucleotides sequences are
XX      CC      useful for detecting cdp's in a sample, for producing a corn ear-specific
XX      CC      profile of gene transcription, for detecting altered gene expression in
XX      CC      hybrid or hybrid plants, and for screening several molecules for specific
XX      CC      binding to the polynucleotide. The cdp's are useful to identify, isolate,

```

```

CC      or extend identical or related corn-ear nucleic acid sequences from DNA
CC      libraries, and in nucleic acid amplification or hybridization techniques
CC      to follow the expression of desirable traits through plant breeding
CC      programs. Preferably, the cdp's are used to identify, evaluate, alter, or
CC      follow the inheritance of desirable characteristics associated with growth,
CC      and development, disease resistance, environmental adaptability, quality,
CC      and yield of corn. The cdp's are also useful as molecular markers for
CC      studying inheritance and multigene traits in a plant breeding program.
CC      The cdp's are useful for producing purified corn-ear polypeptides by
CC      recombinant techniques. They are also useful in diagnostic assays to
CC      detect or confirm conditions or diseases associated with abnormal levels
CC      of cdp expression. ABX81541-ABX89140 represent corn ear-derived
CC      polynucleotides (cdps) of the invention. Note: The sequence data for this
CC      patent did not form part of the printed specification, but was obtained
CC      in electronic format directly from the USPTO web site at
CC      seqdata.uspto.gov/patseq/identity.html
XX      SQ      Sequence 272 BP; 79 A; 52 C; 68 G; 67 T; 0 U; 6 Other;

Query Match      2.3%; Score 74.4; DB 7; Length 272;
Best Local Similarity 58.7%; Pred. No. 7.8e-08;
Matches 142; Conservative 0; Mismatches 97; Indels 3; Gaps 1;

Qy      1777 TATGAGGCTCTTAGAGCAGCGCATTCCTATGCTCCACAGGCGCATCGTGAGTAGTGT 1836
Db      2 TATGAGCAGATMAAGACGATGCTATGCTCCCTGCGGACCGGTGATGAGCGTN 61
Qy      1837 CTGATGTTTGAAGACGACGCTGCTATTTGAGAGCCGCAACGCTTCACCGGAGATT 1896
Db      62 TTATATTTTGAAGAGCTGAGCTGAGGCTCAACATGAGACGCTGCAATMAACATTT 121
Qy      1897 GCTGAGATGGGGTTAGATAGATGCTCTGG--GCTGAGAGCGCAGATGTTTCTGGA 1953
Db      122 GTTANCAAGGTACAGACAGAGATTCATGGCACTTACGANGTTCGATTTGCTCGGT 181
Qy      1954 GGTGTTGCGCACTGTATGCTTCTTCAACAGAGAGATCTGACATATTCATCA 2013
Db      182 GCGAAGAGGCGCATATGCTTCTTCAACAGAGAGATGAGCGATTTAAACAG 241
Qy      2014 CA 2015
Db      242 CA 243

RESULT 4
ABX82003
ID      ABX82003 standard; cDNA; 274 BP.
XX      AC      ABX82003;
XX      DT      24-APR-2003 (first entry)
XX      DE      Corn ear-derived polynucleotide (cpd) #463.
XX      KW      Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;
XX      KW      structural gene; functional gene; regulatory gene;
XX      KW      corn ear-specific profile; gene transcription; gene expression;
XX      KW      hybrid plant; desirable trait expression; plant breeding program;
XX      KW      inheritance; desirable characteristic; growth; development;
XX      KW      disease resistance; environmental adaptability; quality; yield;
XX      KW      multigene trait; plant; gene; ss.
XX      OS      Zea mays.
XX      PN      US6476212-B1.
XX      PD      05-NOV-2002.
XX      PF      14-MAY-1999; 99US-00313294.
XX      PR      26-MAY-1998; 98US-0086722P.
XX      PA      (INCY-) INCYTE GENOMICS INC.

```

XX Lalgudi RV, Ito LY, Sherman BK;
 XX WPI; 2003-208840/20.
 XX
 PT Novel purified corn-ear derived polynucleotide useful as hybridization
 PT probe for detecting polynucleotide in sample, and for identifying,
 PT evaluating, and altering desired characteristics associated with growth,
 PT development.
 XX
 XX Example; SEQ ID NO 463; 390bp; English.
 XX
 XX The present invention relates to the isolation of corn ear-derived
 CC polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022
 CC and SATMON023. Some of the cdps uniquely identify structural, functional,
 CC and regulatory genes of corn ear. The polynucleotide sequences are
 CC useful for detecting cdps in a sample, for producing a corn ear-specific
 CC profile of gene transcription, for detecting altered gene expression in
 CC inbred or hybrid plants, and for screening several molecules for specific
 CC binding to the polynucleotide. The cdps are useful to identify, isolate,
 CC or extend identical or related corn-ear nucleic acid sequences from DNA
 CC libraries, and in nucleic acid amplification or hybridisation techniques
 CC to follow the expression of desirable traits through plant breeding
 CC programs. Preferably, the cdps are used to identify, evaluate, alter, or
 CC follow the inheritance of desired characteristics associated with growth,
 CC and development, disease resistance, environmental adaptability, quality,
 CC and yield of corn. The cdps are also useful as molecular markers for
 CC studying inheritance and multigene traits in a plant breeding program.
 CC The cdps are useful for producing purified corn-ear polypeptides by
 CC recombinant techniques. They are also useful in diagnostic assays to
 CC detect or confirm conditions or diseases associated with abnormal levels
 CC of cdp expression. ABX81541-ABX89140 represent corn ear-derived
 CC polynucleotides (cdps) of the invention. Note: The sequence data for this
 CC patent did not form part of the printed specification, but was obtained
 CC in electronic format directly from the USPTO web site at
 CC seqdata.uspto.gov/psipdsIDentry.html
 XX
 XX Sequence 274 BP; 83 A; 49 C; 73 G; 67 T; 0 U; 2 Other;
 SQ
 Query Match 2.2%; Score 72.2; DB 7; Length 274;
 Best Local Similarity 60.6%; Pred. No. 2.7e-07;
 Matches 152; Conservative 0; Mismatches 95; Indels 4; Gaps 2;
 QY 1777 TATGAGCTCTTAGAGCAGCAGCATTCCTATGCT-CCACAGGCGCATCGGAGATGAGT 1835
 DB 2 TATGAGCAGAGTAAAGCTAGCTATGCTATGCTCTCGGCGCACCGTGAATAGAGCT 61
 QY 1836 TCGATGTTTGAAGAGAGCAGCTGAGCTATTTGAGGCGGACGCTCCACCGGAGTT 1895
 DB 62 GTTATATTTTGAAGAGCTAGCTGAGGCTACATGAGAGCTGAACGCTGCATMAACACTT 121
 QY 1896 AGCTGAGATGGGTTAGATGATTAATTCCTGG---GGTCAGAAAGCGCAGTATGTTCTGG 1952
 DB 122 TGTATATCAAGGATACAGACAGAAATTCATGCACTACCAAGGTTCGATTTGTGCTGG 181
 QY 1953 AGGTGTTCCCAACGTATGAGCTCTCTTCCCAAGAAAGCAAGTCTGACATATTCATCA 2012
 DB 182 TGGGAAAGGCACTAATATGTTTCTTACCAAAAGAGATATGAGGCAATTAACA 241
 QY 2013 ACACCTCTCAAG 2023
 DB 242 GCATTGCCAAG 252
 XX
 XX RESULT 5
 XX ABL33697 standard; DNA; 6668 BP.
 XX ABL33697;
 XX
 XX 26-MAR-2002 (first entry)
 XX
 XX Human immune system associated gene SEQ ID NO: 1670.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostratic; nocrotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antineumatic; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KW ds.
 XX
 XX Homo sapiens.
 OS
 XX
 XX NC0200200928-A2.
 PN
 XX
 XX 03-JUN-2002.
 PD
 XX
 XX 02-JUL-2001; 2001WO-EP007537.
 PF
 XX
 XX 30-JUN-2000; 2000DE-01032529.
 PR
 XX 01-SEP-2000; 2000DE-01043826.
 PR
 XX (EPIC-) EPIDENOMICS AG.
 PA
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX
 XX WPI; 2002-130909/17.
 DR
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PT
 XX
 XX Claim 1; SEQ ID NO 1670; 32pp + Sequence Listing; German.
 PS
 XX
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 CC
 XX
 XX Sequence 6668 BP; 1329 A; 328 C; 1736 G; 3274 T; 0 U; 1 Other;
 SQ
 Query Match 2.1%; Score 68.9; DB 6; Length 6668;
 Best Local Similarity 45.8%; Pred. No. 7.6e-06;
 Matches 238; Conservative 0; Mismatches 282; Indels 0; Gaps 0;
 QY 156 GTTTTAAAGCTTAATAGCTTCTCTATGCTCTCTTCTGCTAGTTATTTCTTCCTC 215
 DB 2742 GTTTTATTTGTTT 2801
 QY 216 CGAGTCTGACACACCTGACCTGCTCCGAGGCTTAAAGCTTACCTCCGTCGTT 275
 DB 2802 TTTTATTTTGTGTTT 2861
 QY 276 ACTCTGAAGTTTCTGCTTGAAGCCTCGAGTGCCTCACCGAGCATTTCTGCTCG 335
 DB 2862 TTTTATTTTGTGTTT 2921
 QY 336 ATTCTCTTTTCTGCTGGAAGAAATTTGCCCTAATGCTCGAATCGAAGTTTGT 395
 DB 2922 TTTTGTT 2981
 QY 396 GCTATGAGTTACTTTTCCCTATATTTTATGTTCTTAGTAAAGATCTGCGTCTA 455
 DB 2962 TTTTATTTGTTT 3041
 QY 456 CTGTTTGTTCATTTGTTGTTGCTTTCACCGTTTACGCTGATTCGAGATTGACTG 515
 DB 3042 TTTTATTTTGTGTTT 3101
 QY 516 TGAAAAATCCCTTCGTTTGTGTTTGTCTTCAATAAATGAGATGATACCTTTTGT 575

265 C T C C G T C G T T A C T C T G A G T T T C T G C C T T A G A G C C T T C C G A T C G C C T C A C C G C A T G C A 324

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

Query Match 2.1%; Score 68.6; DB 5; Length 556;
Best Local Similarity 48.3%; Pred. No. 2.8e-06;
Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
265 CTCGCGCTTACTCTGTAAGTTTCTGCTTAGAGCCCTCGATGCGCTACCGCATGCA 324
466 CMCCTCTCCCCCTTTTATATTTTATTTTTCCTCCCTCCCTCCCATTAATA 427
335 TTCTGTGCTGATTTCTCTTTCTCTGCTGGAATAATGCGCTATGCTCGATTTG 384
426 TTTTATTTTATTTTATTTTATTAATAAATTAACNCCCTTTTATTTTATTTT 367
365 AAGTTTGTGCTAGGCTTACTTTTCCCTATTTTATTAAGTTCTTAGTAACATA 444
366 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 307
445 CCGGCTTACTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 504
306 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 247
505 GATTTGACTGGAATAAATCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 564
246 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 187
565 CTACCTTTGCTGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 624
186 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 127
625 TCACGTTGATGTTGATTTTGAATTTT 653
126 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 98
```

RESULT 8

ABV42105/c

ID ABV42105 standard; cDNA; 556 BP.

XX AC ABV42105;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 42096.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX PD 23-AUG-2001.

XX PF 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219007P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JB;

XX DR WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX PT prostate cells and correlating with presence of prostate cancer, useful
XX PT for detecting presence of prostate cancer, stage of prostate cancer.

XX PS Claim 1; Page 8441; 11750pp; English.

XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX CC specification or its complement. (I) is useful for: (a) assessing whether
XX CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
XX CC progression of prostate cancer in a patient; (c) assessing the efficacy
XX CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX CC (e) selecting a composition for inhibiting prostate cancer in a patient;
XX CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX CC determining whether prostate cancer has metastasized in a patient; (h)
XX CC assessing the aggressiveness or indolence of prostate cancer in a patient
XX CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;

Query Match 2.1%; Score 68.6; DB 5; Length 556;
Best Local Similarity 48.3%; Pred. No. 2.8e-06;
Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

```
265 CTCGCGCTTACTCTGTAAGTTTCTGCTTAGAGCCCTCGATGCGCTACCGCATGCA 324
486 CMCCTCTCCCCCTTTTATATTTTATTTTTCCTCCCTCCCTCCCATTAATA 427
325 TTCTGTGCTGATTTCTCTTTCTCTGCTGGAATAATGCGCTATGCTCGATTTG 384
426 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 367
385 AAGTTTGTGCTAGGCTTACTTTTCCCTATTTTATTAAGTTCTTAGTAACATA 444
366 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 307
445 CCGGCTTACTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 504
306 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 247
505 GATTTGACTGGAATAAATCTGCTGTTTGTGTTGTTGTTGTTGTTGTTGTTGTT 564
246 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 187
565 CTACCTTTGCTGTTGATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 624
186 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 127
625 TCACGTTGATGTTGATTTTGAATTTT 653
126 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 98
```

RESULT 9

ABV43601/c

ID ABV43601 standard; cDNA; 556 BP.

XX AC ABV43601;

XX DT 16-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 43592.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN WO200160860-A2.

XX 23-AUG-2001.
PD 20-FEB-2001; 2001WO-US0005171.
PF 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA Schlegel R, Endege WO, Monahan JE;
PI WPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 8681-8682; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV6213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC assessing whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 556 BP; 389 A; 1 C; 51 G; 113 T; 0 U; 2 Other;
Query Match 2.1%; Score 68.6; DB 5; Length 556;
Best Local Similarity 48.3%; Pred. No. 2.8e-06;
Matches 188; Conservative 0; Mismatches 201; Indels 0; Gaps 0;
QY 265 CTCGCGTGTACCTGTAAGTTTCTGCTTAGAGCGCTCGATGCGCTACCGCATGCA 324
DB 486 CNGCCTCTCCCTTTTAATTAATTTTATTTTCTCCCCCTCCCAATAATA 427
QY 325 TTCTGCTGATTTCTTTTCTGCTGGAATAATGCTTAATGTTCTCGATTTCG 384
DB 426 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 367
QY 385 AAGGTTTGTGCTAGAGGTTACTTTTCCCTAATTTTAAAGTTCTAGAGCAATA 444
DB 366 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 307
QY 445 CTTGCGCTACTGTTTGTGCTATTTGTTGCTTCAACCGTTAGTGGCTAGCGGA 504
DB 306 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 247
QY 505 GATTTGACTGGAATAATCTTCGTTTGTGTTTGTGTTTCAATATAATCGGATTGAT 564
DB 246 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 187
QY 565 CTACCTTTTGTGCTGATGTTGTTTGTGAGCCATAGCGTTGGCTGTTAATACT 624
DB 186 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 127
QY 625 TCACGTCATGCTGATTTGAGATT 653
DB 126 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 98
RESULT 10

AA545347
ID AA545347 standard; DNA; 9539 BP.
XX
AC AA545347;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated complementary DNA associated with cell cycle #26.
XX
KW Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
OS Homo sapiens.
XX
XX WO200168911-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP002945.
XX
XX 15-MAR-2000; 2000DE-01013847.
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX (EPig-) EPiGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2001-602751/68.
XX
XX Designing primers and probes for analyzing diseases associated with
PT cytosine methylation state e.g. arthritis, cancer, aging,
PT arteriosclerosis comprising fragments of chemically modified genes
PT associated with cell cycle.
XX
PS Claim 1; SEQ ID NO 52; 28pp; English.
XX
CC Sequences AA545296-AA545520 represent chemically pretreated genomic DNA
CC molecules associated with the cell cycle and specific PCR primers of the
CC invention. The sequences are useful for detecting the methylation state
CC of all CpG dinucleotides in a sequence and therefore for analyzing
CC associated diseases. By analysing cytosine methylations in the pretreated
CC DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC of existing diseases or the predisposition to specific diseases can be
CC ascertained. The parameters may be compared to another set of genetic
CC and/or epigenetic parameters, the differences serving as basis for
CC diagnosis and/or prognosis events which are disadvantageous to patients.
CC The sequences of the invention are useful for the diagnosis and therapy
CC of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC aging, glomerular disease, Lewy body disease, arthritis,
CC arteriosclerosis, solid tumours and cancers
XX
SQ Sequence 9539 BP; 2078 A; 121 C; 2281 G; 5059 T; 0 U; 0 Other;
Query Match 2.0%; Score 66.8; DB 4; Length 9539;
Best Local Similarity 46.7%; Pred. No. 2.7e-05;
Matches 212; Conservative 0; Mismatches 242; Indels 0; Gaps 0;
QY 251 TTTAACTTACGTTCTCCGTTTACTCTGTAAGTTTCTGCTAGAGCTCCGATCG 310
DB 202 TTTTCTTTTGTATTTTTCGTTTATTTTGTGTTTCTTTTCTTTTCTTTTCTTTTCTTT 261
QY 311 CCTACCGCATGATTTGCTGATTTCTCTTTTCTTCTGCTGGAATAATGCCCTAA 370
DB 262 GTTTTCTTTTGTGCTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 321
QY 371 TGTTCGATTTGGAAGTTTGTGCTAGAGGTTACTTTTCCCTAATTTTAAAGTT 430

KW Nijmegen breakage syndrome; Werner syndrome; immunodeficiency;
XX trichiodystrophy; Fanconi's anaemia; solid tumour; cancer; ds.
OS unidentified.
XX WO200181622-A2.
PN
XX
PD 01-NOV-2001.
XX
PF 06-APR-2001; 2001WO-EP003972.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-034446/04.
XX
XX New nucleic acid derived from genes associated with DNA repair, useful
PT for diagnosis, e.g. of ataxia telangiectasia, by determination of
PT cytosine methylation.
XX
PS Claim 1; SEQ ID NO 122; 25pp + Sequence Listing; English.
XX
XX The invention relates to nucleic acids containing a sequence of at least
CC 18 nucleotides of chemically treated DNA of genes associated with DNA
CC repair, and their complements. The invention also relates to nucleic
CC acids comprising at least 18 base pairs of the chemically pretreated DNA
CC of genes associated with DNA repair selected from PMS2L1, PMS2L12,
CC PMS2L2, PMS2L3, PMS2, L4, PMS2L5, MGMT, MSH2, NUDT1, TDB, INPPL1,
CC RFC4, DDT1L, FANCB, or XRCC8. Nucleic acids of the invention and related
CC oligomers, are useful for diagnosis of diseases associated with gene
CC repair, specifically ataxia telangiectasia, aging, Bloom's syndrome,
CC Cockayne syndrome, Nijmegen breakage syndrome or Werner syndrome,
CC immunodeficiency, trichiodystrophy, Fanconi's anaemia, solid tumours
CC and cancer, particularly by determining status of cytosine methylation
CC and/or by detecting single-nucleotide polymorphisms. Determination of
CC individual methylation patterns may allow development of individualised
CC therapies. The sequences given in records ABL92192-ABL92335 represent
CC chemically pre-treated DNA fragments from genes associated with DNA
CC repair, and their complements. Note: The sequence data for this patent is
CC not represented in the specification, but is based on sequence
CC information supplied by the European Patent Office
XX
SQ Sequence 8079 BP; 1992 A; 212 C; 1962 G; 3913 T; 0 U; 0 Other;
Query Match 2.0%; Score 66.4; DB 6; Length 8079;
Best Local Similarity 48.2%; Pred. No. 3.2e-05;
Matches 187; Conservative 0; Mismatches 201; Indels 0; Gaps 0;
QY 333 TCGATTCCTCTTTCTTCGCGGAAAAATGCGCTATGTTTCGATTTGGAAGTTT 392
Db 5327 TTGTTGATGATTTTATATATATAGAGAGTGCTTATGTTGATTTTGTGTTTTT 5386
QY 393 TGGCTAAGGCTACTTTTCCCTATATTTTATGTTCTTAGGTAACGATACCTGCGTC 452
Db 5387 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5446
QY 453 TTACTGTTTGTCTATTTGTTGCTTACCGCTTAGCGCTGATCGAGATTTGA 512
Db 5447 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5506
QY 513 CTGTGAAAAATCCCTGCTTTTGTGTTTGTTCATATAAATGAGATGATACCTTT 572
Db 5507 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5566
QY 573 TGGCTTGTGATGTTGTTTGTGAGCCTAGCGTTGCTGTATATCTACGCTTC 632
Db 5567 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5626

QY 633 ATGTGTGATTTTGAGATTGTTGATGACTGTGGGTTTCTTGATGCTATAGTTGTA 692
Db 5627 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5686
QY 693 AAAATGATTTCTAGGCTGTGCTCAATG 720
Db 5687 GTGTAGTGTATGTTATGTTATGTTATTTG 5714
RESULT 13
ID AAX33181/c
XX AAX33181 standard; DNA; 6644 BP.
XX
XX AAX33181;
AC
XX
XX 25-JUN-1999 (first entry)
DT
XX
XX Base sequence of the plasmid pRX-lres-bar.
DE
XX
XX Cowpox virus; bar; viral vector; expression; apoptosis; resistance; crmA;
KW bcl-2; bcl-xl; FLIP; survivin; IAP; ILP; adenovirus; cancer;
KW autoimmune disease; graft rejection reaction; inflammation;
KW inflammatory disease; ss.
XX
XX
OS Synthetic.
OS Cowpox virus.
XX
XX WO9913073-A2.
XX
XX 18-MAR-1999.
PD
XX
XX 07-SEP-1998; 98WO-JP004010.
PR
XX
XX 08-SEP-1997; 97JP-00259235.
PR
XX
XX (RPRG-) RPR GENCELL ASIA PACIFIC INC.
PA
XX
XX Hamada H;
PI
XX
XX WPI; 1999-243728/20.
DR
XX
XX
PT
XX
XX New apoptosis-resistant virus-sensitive cell.
PS
XX
XX Example 1; Page 38-41; 51pp; English.
XX
XX The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRX-lres-bar, which contains the cowpox virus bar gene, and is used in an
CC example from the present invention
XX
SQ Sequence 6644 BP; 2166 A; 1573 C; 1424 G; 1481 T; 0 U; 0 Other;
Query Match 2.0%; Score 66; DB 2; Length 6644;
Best Local Similarity 45.4%; Pred. No. 3.6e-05;
Matches 237; Conservative 0; Mismatches 285; Indels 0; Gaps 0;

XX The present invention describes an apoptosis-resistant virus-sensitive
CC cell line into which an apoptosis resistance gene has been introduced.
CC The recombinant viruses generated are capable of expressing apoptosis-
CC associated genes. These can then be used in a variety of diseases for
CC which the induction of apoptosis by gene transfer, or where the
CC inhibition of harmful apoptosis, is therapeutic. The recombinant viruses
CC are useful as vectors for gene therapy which can be applied to cancer
CC therapy for destroying cancer cells selectively, the treatment of
CC autoimmune diseases and graft rejection reaction, and apoptosis induction
CC therapy for inflammatory cells in inflammatory diseases. Prior arts have
CC encountered the problem where if an adenovirus vector capable of
CC expressing an apoptosis-associated gene is introduced into animal cells,
CC the cells producing the virus will be destroyed because the period of
CC time required to induce cell death by apoptosis is shorter than that
CC required to replicate and produce the virus, resulting in failure to
CC obtain a recombinant virus having the integrated apoptosis-associated
CC gene. In this invention an apoptosis-resistant 293 cell line (having an
CC apoptosis resistant gene introduced) is established and overcomes the
CC problem. The present sequence represents the base sequence of the plasmid
CC pRX-Bcl-xl-bar, which contains the human Bcl-xl gene, and is used in an
CC example from the present invention
XX

SQ Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 U; 0 Other;

Sequence 7372 BP; 2353 A; 1749 C; 1649 G; 1621 T; 0 U; 0 Other;

```
Query Match      2.0%; Score 66; DB 2; Length 7372;
Best Local Similarity: 45.4%
Best %: 3.0%
```

Best Local Similarity	45.4%;	Pred. No. 3.8e-05;
Matches 237; Conservative	0;	Mismatches 285; Indels 0;
		Gaps 0

AC AAX33180,
XX
DT 25-JUN-1999 (first entry)

This Page Blank (uspto)


```
US-10-311-455-1214
; Sequence 1214, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 1214
; LENGTH: 9888
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1214

Query Match
Best Local Similarity 100.0%; Score 22; DB 14; Length 9888;
Pred. No. 3.8; Mismatches 0; Indels 0; Gaps 0;
Matches 22; Conservative 0;

QY 3142 TAGATTTTGGTTTGGT 3163
DB 8219 TAGATTTTGGTTTGGT 8240

RESULT 3
US-10-027-632-250839
; Sequence 250839, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; PRIOR FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 250839
; LENGTH: 2164
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-250839

Query Match
Best Local Similarity 100.0%; Score 21; DB 15; Length 2164;
Pred. No. 11; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 3154 TTTTGGTAAATTTCATA 3174
DB 842 TTTTGGTAAATTTCATA 862

RESULT 4
US-10-311-455-482
; Sequence 482, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 482
; LENGTH: 6048
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-482

Query Match
Best Local Similarity 100.0%; Score 21; DB 14; Length 6048;
Pred. No. 12; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 526 TTGGTTTGGTTTGGTT 546
DB 5395 TTGGTTTGGTTTGGTT 5415

RESULT 5
US-10-311-455-4
; Sequence 4, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determining Cytosine Methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; PRIOR FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 4
; LENGTH: 6072
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-4

Query Match
Best Local Similarity 100.0%; Score 21; DB 14; Length 6072;
Pred. No. 12; Mismatches 0; Indels 0; Gaps 0;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```


Best Local Similarity 100.0%; Pred. No. 32,
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3164 AAAATTCATATGAAGCTTA 3183
DB 153 AAAATTCATATGAAGCTTA 134

RESULT 10

US-10-074-475-133
; Sequence 133, Application US/10074475
; Publication No. US2003092898A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Reardon, Heve
; APPLICANT: Karta, Kalpana
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
; FILE REFERENCE: DEX-0313
; CURRENT APPLICATION NUMBER: US/10/074,475
; PRIOR FILING DATE: 2002-02-13
; PRIOR APPLICATION NUMBER: 60/268,292
; NUMBER OF SEQ ID NOS: 295
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 133
; LENGTH: 1023
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-074-475-133

Query Match
Best Local Similarity 100.0%; Score 20; DB 14; Length 1023;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3136 TTTCCTAGTATTTGTTT 3155
DB 832 TTTCCTAGTATTTGTTT 851

RESULT 11

US-10-424-599-91177
; Sequence 91177, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO: 91177
; LENGTH: 1377
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_53343C.1
US-10-424-599-91177

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 1377;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1368 CAGTGCATTGTCCAGCTTG 1387

DB 359 CAGTGCATTGTCCAGCTTG 378

RESULT 12

US-10-282-122A-10949/c
; Sequence 10949, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeelbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EPIRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 10949
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
US-10-282-122A-10949

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 1461;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2047 ATATATGCTTTTGTGTTGT 2066
DB 1073 ATATATGCTTTTGTGTTGT 1054

RESULT 13

US-10-424-599-91176
; Sequence 91176, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

```

? TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
? FILE REFERENCE: 38-21 (532223) B
? CURRENT APPLICATION NUMBER: US/10/424,559
? CURRENT FILING DATE: 2003-04-28
? NUMBER OF SEQ ID NOS: 285684
? SEQ ID NO 91176
? LENGTH: 1589
? TYPE: DNA
? ORGANISM: Glycine max
? FEATURE:
? OTHER INFORMATION: Clone ID: PAT_MRT3847_53342C.1
? US-10-424-559-91176

```

Query Match	0.6%;	Score 20;	DB 12;	Length 1589;
Best Local Similarity	100.0%;	Pred. No. 35;		
Matches 20;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy		1368	CAGTGGCATTGTCCAGCTTG	1387
Db		359	CAGTGGCATTGTCCAGCTTG	378

```

RESULT 14
US-10-108-260A-1756
; Sequence 1756, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: HI-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1756
; LENGTH: 2418
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-1756

```

```
Query Match      0.6%; Score 20; DB 15; Length 2418;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3165 AAATTTCATATGAAAGTTAG   3184  
        ||| | | | | | | | | |  
Db     1988 AAATTTCATATGAAAGTTAG   2007
```

RESULT 15

US-09-947-953-1/c
; Sequence 1, Application US/09947953
; Patent No. US20020155101A1

; GENERAL INFORMATION:

APPLICANT: DONAHUE, J. KEVIN

APPLICANT: MARRBAN, EDUARDO

FILE REFERENCE: 31600/EC415

FILE REFERENCE: 1699/56413
CURRENT ABILICATION NUMBER: HS/08/047 053

COMPONENT: 3001-08-06

PRIOR APPLICATION NUMBER: 60/230 311

PRIOR FILING DATE: 2001-09-06

PRIOR APPLICATION NUMBER: 60/295.889

PRIOR FILING DATE: 2001-06-05

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn Ver. 2.1

: SEQ ID NO 1

```
; LENGTH: 2985
```

```

; TYPE: DNA

```

ORGANISM: Homo sapiens

US-09-947-953-1

Query Match 0.6%; Score 20; DB 9; Length 2985;

Best Local Similarity 100.0%; Pred. NO. 38;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	71	CGAAAAAGGAAAAAAAAGG	90
Db	25	CGAAAAAGGAAAAAAAAGG	6

Search completed: March 30, 2004, 23:20:26
Job time : 2047.85 secs

This Page Blank (uspto)

TYPE: DNA
ORGANISM: Escherichia coli
US-09-736-734-1

Query Match 0.6%; Score 20; DB 4; Length 654;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3164 AAAATTCATATGAAAGTTA 3183
Db 153 AAAATTCATATGAAAGTTA 134

RESULT 3
US-09-615-192A-109/c
Sequence 109, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003c4U
CURRENT FILING DATE: 2000-07-12
PRIOR FILING DATE: 1997-11-21
PRIOR APPLICATION NUMBER: US 08/975,316
PRIOR FILING DATE: 1996-09-11
PRIOR APPLICATION NUMBER: US 09/169,789
PRIOR FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 405
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 109
LENGTH: 481
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-615-192A-109

Query Match 0.6%; Score 19; DB 4; Length 481;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 TGGTTCAGGTTTGGCAGG 787
Db 470 TGGTTCAGGTTTGGCAGG 452

RESULT 4
US-09-169-789-109/c
Sequence 109, Application US/09169789
Patent No. 6653528
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003c2
CURRENT FILING DATE: US/09/169,789
EARLIER FILING DATE: 1998-10-09
EARLIER APPLICATION NUMBER: US 08/975,316
EARLIER FILING DATE: 1997-11-21
EARLIER APPLICATION NUMBER: US 08/713,000
NUMBER OF SEQ ID NOS: 185
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 109
LENGTH: 481
TYPE: DNA
ORGANISM: Eucalyptus grandis
US-09-169-789-109

Query Match 0.6%; Score 19; DB 4; Length 481;

Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 TGGTTCAGGTTTGGCAGG 787
Db 470 TGGTTCAGGTTTGGCAGG 452

RESULT 5
US-08-975-316-27/c
Sequence 27, Application US/08975316
Patent No. 5952486
GENERAL INFORMATION:
APPLICANT: BLOKSBERG, Leonard N., HAVUKKALA, Ilkka
TITLE OF INVENTION: MATERIALS AND METHODS FOR
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,316
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/713,000
FILING DATE: September 11, 1996
ATTORNEY/AGENT INFORMATION:
NAME: SLEATH, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000/1003C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 495 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-975-316-27

Query Match 0.6%; Score 19; DB 2; Length 495;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 769 TGGTTCAGGTTTGGCAGG 787
Db 484 TGGTTCAGGTTTGGCAGG 466

RESULT 6
US-09-615-192A-27/c
Sequence 27, Application US/09615192A
Patent No. 6410718
GENERAL INFORMATION:
APPLICANT: Bloksberg, Leonard N.
TITLE OF INVENTION: Materials and Methods for the
FILE REFERENCE: 11000.1003c4U
CURRENT APPLICATION NUMBER: US/09/615,192A

;; CURRENT FILING DATE: 2000-07-12
;; PRIOR APPLICATION NUMBER: US 08/975,316
;; PRIOR FILING DATE: 1997-11-21
;; PRIOR APPLICATION NUMBER: US 08/713,000
;; PRIOR FILING DATE: 1996-09-11
;; PRIOR APPLICATION NUMBER: US 09/169,789
;; PRIOR FILING DATE: 1998-10-09
;; NUMBER OF SEQ ID NOS: 405
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 27
;; LENGTH: 495
;; TYPE: DNA
;; ORGANISM: Eucalyptus grandis
US-09-615-192A-27

Query Match 0.6%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAAGTTTGGCAGG 787
|||
Db 484 TGGTTCAAGTTTGGCAGG 466

RESULT 7
US-09-169-789-27/c
;; Sequence 27, Application US/09169789
;; Patent No. 6653528
;; GENERAL INFORMATION:
;; APPLICANT: Bloksberg, Leonard N.
;; APPLICANT: Havukkala, Ilkka
;; TITLE OF INVENTION: Materials and Methods for the
;; FILE REFERENCE: 11000.1003c2
;; CURRENT APPLICATION NUMBER: US/09/169,789
;; CURRENT FILING DATE: 1998-10-09
;; EARLIER APPLICATION NUMBER: US 08/975,316
;; EARLIER FILING DATE: 1997-11-21
;; EARLIER APPLICATION NUMBER: US 08/713,000
;; EARLIER FILING DATE: 1996-09-11
;; NUMBER OF SEQ ID NOS: 185
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 27
;; LENGTH: 495
;; TYPE: DNA
;; ORGANISM: Eucalyptus grandis
US-09-169-789-27

Query Match 0.6%; Score 19; DB 4; Length 495;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 769 TGGTTCAAGTTTGGCAGG 787
|||
Db 484 TGGTTCAAGTTTGGCAGG 466

RESULT 8
US-08-484-105-19
;; Sequence 19, Application US/08484105
;; Patent No. 5589341
;; GENERAL INFORMATION:
;; APPLICANT: STILLMAN, Bruce
;; APPLICANT: BEIL, Stephen P
;; APPLICANT: KOBAYASHI, Ryuji
;; APPLICANT: RINE, Jasper
;; APPLICANT: FOSS, Margit
;; APPLICANT: MCNALLY, Francis J
;; APPLICANT: LAURENSEN, Patricia
;; APPLICANT: HERSKOWITZ, Ira
;; APPLICANT: LI, Joachim J
;; APPLICANT: GAVIN, Kimberly
;; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESS: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;; STREET: 4 Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-4187
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patencin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/484,105
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Osman Ph.D., Richard Aron
;; REGISTRATION NUMBER: 36,627
;; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 494-8700
;; TELEFAX: (415) 494-8771
;; TELE: 910 277299
;; INFORMATION FOR SEQ ID NO: 19:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1480 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 277..1365
US-08-484-105-19

Query Match 0.6%; Score 19; DB 1; Length 1480;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1655 TAAGTGAATTCCTCTGT 1673
|||
Db 238 TAAGTGAATTCCTCTGT 256

RESULT 9
US-08-484-106-19
;; Sequence 19, Application US/08484106
;; Patent No. 5614618
;; GENERAL INFORMATION:
;; APPLICANT: STILLMAN, Bruce
;; APPLICANT: BEIL, Stephen P
;; APPLICANT: KOBAYASHI, Ryuji
;; APPLICANT: RINE, Jasper
;; APPLICANT: FOSS, Margit
;; APPLICANT: MCNALLY, Francis J
;; APPLICANT: LAURENSEN, Patricia
;; APPLICANT: HERSKOWITZ, Ira
;; APPLICANT: LI, Joachim J
;; APPLICANT: GAVIN, Kimberly
;; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES
;; NUMBER OF SEQUENCES: 24
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
;; STREET: 4 Embarcadero Center, Suite 3400
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-4187
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible

RESULT 13
 US-09-595-684B-30
 Sequence 30, Application US/09595684B
 Patent No. 6544766
 GENERAL INFORMATION:
 APPLICANT: Bersud, Christophe
 APPLICANT: Ohashi, Cara
 APPLICANT: Sakowicz, Roman
 APPLICANT: Vaisberg, Eugene
 APPLICANT: Wood, Kenneth
 APPLICANT: Yu, Ming
 TITLE OF INVENTION: Human kinases and methods of producing
 TITLE OF INVENTION: and purifying human kinases
 FILE REFERENCE: cytopo36
 CURRENT APPLICATION NUMBER: US/09/595,684B


```
; CURRENT FILING DATE: 2002-06-24
; PRIOR APPLICATION NUMBER: 09/295,612
; PRIOR FILING DATE: 2000-04-20
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FASTSEQ For Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 8257
; TYPE: DNA
; ORGANISM: Human
US-09-595-684B-30
```

```
Query Match 0.6%; Score 19; DB 4; Length 8257;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2873 TTGATCTGAGAGAAAGATT 2891
DB 2141 TTGATCTGAGAGAAAGATT 2159
```

RESULT 14

```
US-09-620-312D-130
; Sequence 130, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yundong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662e1 Nucleic Acids and
; FILE REFERENCE: 784C1P2B
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_fl_genes Version 1.0
; SEQ ID NO 130
; LENGTH: 8503
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (91)..(8082)
US-09-620-312D-130
```

```
Query Match 0.6%; Score 19; DB 4; Length 8503;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2873 TTGATCTGAGAGAAAGATT 2891
DB 2141 TTGATCTGAGAGAAAGATT 2159
```

```
RESULT 15
US-09-163-748C-3/c
```

```
; Sequence 3, Application US/09163748C
; Patent No. 6509172
; GENERAL INFORMATION:
; APPLICANT: Debacker, Oliver
; APPLICANT: Van den Eynde, Benoit
; APPLICANT: Boon-Falleur, Thierry
; TITLE OF INVENTION: Isolated Nucleic Acid Molecules Which Encode Member Of The Gage
; FILE REFERENCE: IUD 5558
; CURRENT APPLICATION NUMBER: US/09/163,748C
; CURRENT FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 27
; SEQ ID NO 3
; LENGTH: 9531
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 92,232,1041,7412,9038-9040
; OTHER INFORMATION: Identity of several nucleotides not known
US-09-163-748C-3
```

```
Query Match 0.6%; Score 19; DB 4; Length 9531;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 91 TACAAAATGAAAAACAAA 109
DB 7045 TACAAAATGAAAAACAAA 7027
```

```
Search completed: March 30, 2004, 18:11:43
Job time : 250.688 secs
```

This Page Blank (uspto)

XX Beclin C, Elmayer T, Vaucheret H;
 XX WPI; 2001-159529/16.
 DR New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
 PT resistance in plants and, when inhibited, for increasing transgene
 PT expression.
 XX Claim 1; Page 31-32; 36pp; French.
 PS
 CC The present sequence represents the genomic sequence of the Arabidopsis
 CC thaliana SGS3 gene. The SGS3 gene is essential for post-transcriptional
 CC inactivation (degradation of RNA) and for resistance to viruses.
 CC Overexpression of SGS3 results in plants with increased resistance to
 CC viruses; while inactivation of SGS3 in transgenic plants (e.g. by
 CC expressing antisense RNA, by mutation or by homologous recombination)
 CC increases the level of the transgene product. This product may e.g.
 CC impart resistance (to herbicide, insects or pathogens), alter contents of
 CC essential fatty acids or proteins, or is pharmaceutically active, e.g. an
 CC immunoglobulin or interferon
 XX
 SQ Sequence 3275 BP; 956 A; 561 C; 804 G; 954 T; 0 U; 0 Other;
 Query Match 100.0%; Score 3275; DB 4; Length 3275;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAACAAACAAATTTAGCAATGCTTTGTCGATTAATTAATAGTGGACAA 60
 DB 1 GACAAACAAACAAATTTAGCAATGCTTTGTCGATTAATTAATAGTGGACAA 60
 QY 61 TTAAGTTAAGCAAAAGAAAGAAAGTACAAATGAAACAAATCAACTGAT 120
 DB 61 TTAAGTTAAGCAAAAGAAAGAAAGTACAAATGAAACAAATCAACTGAT 120
 QY 121 GAAATTTGGAGTCCAGAAATCGAAAGAGCGCTTTAGAGCTTAATAGCTTC 180
 DB 121 GAAATTTGGAGTCCAGAAATCGAAAGAGCGCTTTAGAGCTTAATAGCTTC 180
 QY 181 ATTGTCCTCTTGTGAGTTATTTCTTCTCCGAGTCTGACCTCACTCTGAC 240
 DB 181 ATTGTCCTCTTGTGAGTTATTTCTTCTCCGAGTCTGACCTCACTCTGAC 240
 QY 241 TCTCCGAGCTTTAAACTTACGTTCTCCGTCGTTACTCTGAAATTTTCTGCTT 300
 DB 241 TCTCCGAGCTTTAAACTTACGTTCTCCGTCGTTACTCTGAAATTTTCTGCTT 300
 QY 301 CCTCGATCGCTCACCGCATGCAATTCGTGCTGCAATTCCTTTTCTGCTGAAA 360
 DB 301 CCTCGATCGCTCACCGCATGCAATTCGTGCTGCAATTCCTTTTCTGCTGAAA 360
 QY 361 ATGCGCTAATGTTCTGATTCGAAGTTTGTGCGATAGGGTACTTTTCCCTATA 420
 DB 361 ATGCGCTAATGTTCTGATTCGAAGTTTGTGCGATAGGGTACTTTTCCCTATA 420
 QY 421 TTTTAATGTTCTTAAAGTACGATCCTGCTTACTGTTTGTGATTTTGTGCT 480
 DB 421 TTTTAATGTTCTTAAAGTACGATCCTGCTTACTGTTTGTGATTTTGTGCT 480
 QY 481 TTTACCGTTTGTGCTGATGAGGATTTTACCTGTGAAATTCCTTGTGTTTGT 540
 DB 481 TTTACCGTTTGTGCTGATGAGGATTTTACCTGTGAAATTCCTTGTGTTTGT 540
 QY 541 TTGTTTCAATATAATCGAATGATCTTCTTGTGCTTGAATGTTTGTGACCT 600
 DB 541 TTGTTTCAATATAATCGAATGATCTTCTTGTGCTTGAATGTTTGTGACCT 600
 QY 601 ATGCGTTGTTGGCTTTGTAATCACTTCACTGATGATGATTTTGAATTTGTAG 660
 DB 601 ATGCGTTGTTGGCTTTGTAATCACTTCACTGATGATGATTTTGAATTTGTAG 660
 QY 661 ACTGTGGGTTCTTTGTGCTATAGGTTGTAATAATGAGTTCTAGGGCTGTCATGT 720

DB 661 ACTGTGGGTTCTTTGTGCTATAGGTTGTAATAATGAGTTCTAGGGCTGTCATGT 720
 QY 721 CTAAGAAAAGAACGTTCAAGGTGTATATAGGCTGAGTGAACGTTGTTCAAGTT 780
 DB 721 CTAAGAAAAGAACGTTCAAGGTGTATATAGGCTGAGTGAACGTTGTTCAAGTT 780
 QY 781 TGGCAAGGACGACGCTGCTTCTTCAAGATGATGAGAGAGTGGAGTCAATTTCA 840
 DB 781 TGGCAAGGACGACGCTGCTTCTTCAAGATGATGAGAGAGTGGAGTCAATTTCA 840
 QY 841 AGAAGAACAAAGAACAAACGAGAAACCTTCTGAAAACTTGGGTTTCTCAATTCGA 900
 DB 841 AGAAGAACAAAGAACAAACGAGAAACCTTCTGAAAACTTGGGTTTCTCAATTCGA 900
 QY 901 ATCTCTTGAAGCTTGGGTTGTCTAGCAGCAAGGAGGATGACCAAGTATCTGGAGAG 960
 DB 901 ATCTCTTGAAGCTTGGGTTGTCTAGCAGCAAGGAGGATGACCAAGTATCTGGAGAG 960
 QY 961 GAAACAAATGATTCGGAGAGGTACCGCAATGTCGGGCAATTCAAGCTAATATCTG 1020
 DB 961 GAAACAAATGATTCGGAGAGGTACCGCAATGTCGGGCAATTCAAGCTAATATCTG 1020
 QY 1021 GTGGGGAGAGAGGTTGAGCAGAAAGTATGATTAACAATTTGTGCAACCCCACTGTAT 1080
 DB 1021 GTGGGGAGAGAGGTTGAGCAGAAAGTATGATTAACAATTTGTGCAACCCCACTGTAT 1080
 QY 1081 CTGGCTCTCTTTGGAAGAGATGGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGCACA 1140
 DB 1081 CTGGCTCTCTTTGGAAGAGATGGAATTTGGCAGGCAAGAGAGGTTCTGCTCAGCACA 1140
 QY 1141 CAGCTGCGAGAGTTTCTGAGCTGAGAGTATGATGATTAATGCTTCTGAGGAAGAGA 1200
 DB 1141 CAGCTGCGAGAGTTTCTGAGCTGAGAGTATGATGATTAATGCTTCTGAGGAAGAGA 1200
 QY 1201 ATGATTCGATGCTTTGATGATGATTTCTGATGACGACCTTCAAGTATGATGATCTGG 1260
 DB 1201 ATGATTCGATGCTTTGATGATGATTTCTGATGACGACCTTCAAGTATGATGATCTGG 1260
 QY 1261 ATGTAGTCAAAAGACCAATGATATCGAAAGCAGAAATAGTGTCAAAAAAGTCTTTG 1320
 DB 1261 ATGTAGTCAAAAGACCAATGATATCGAAAGCAGAAATAGTGTCAAAAAAGTCTTTG 1320
 QY 1321 GCAGCTTGATATGCTTGTGATGATGATTTCTGATGACGACCTTCAAGTATGATGATCTGG 1380
 DB 1321 GCAGCTTGATATGCTTGTGATGATGATTTCTGATGACGACCTTCAAGTATGATGATCTGG 1380
 QY 1381 CAGCTTGTCAAGAACGACCTGATGATGATTTGATGATGATGATGATGATGATGATGATG 1440
 DB 1381 CAGCTTGTCAAGAACGACCTGATGATGATTTGATGATGATGATGATGATGATGATGATG 1440
 QY 1441 ATGCGAGGACAAAGAGGCTTACGAGGATTAAGCTTCAATGAGATTTGCTGAAATTTAG 1500
 DB 1441 ATGCGAGGACAAAGAGGCTTACGAGGATTAAGCTTCAATGAGATTTGCTGAAATTTAG 1500
 QY 1501 AAAAGGATCTACAGATGAGAGGCAATGATCTTGTGCTGATGATTTTATGAGGAGT 1560
 DB 1501 AAAAGGATCTACAGATGAGAGGCAATGATCTTGTGCTGATGATTTTATGAGGAGT 1560
 QY 1561 GGAAGGTTTGGGTGAGATGAAAAGATTAATGAAATTTGCTGCTCAATGATGATCA 1620
 DB 1561 GGAAGGTTTGGGTGAGATGAAAAGATTAATGAAATTTGCTGCTCAATGATGATCA 1620
 QY 1621 TCAATGATCTAGATGATGATTAAGGACGATTAAGATGAGTGAATTTCTGCTTTTAC 1680
 DB 1621 TCAATGATCTAGATGATGATTAAGGACGATTAAGATGAGTGAATTTCTGCTTTTAC 1680
 QY 1681 TTCTTAATTTTCTCTGATTTTCTGATGATGATTTTGAATGATGATGATGATGATGATGATG 1740
 DB 1681 TTCTTAATTTTCTCTGATTTTCTGATGATGATTTTGAATGATGATGATGATGATGATGATG 1740
 QY 1741 ATGCGAACCAAGAGCTGCTGATATCTTGAACAAATGAGGCTTTAGAGCAGCCAT 1800

Db 1741 ATGGCAACCAAGAGCTGCTGAAATCTTCGACCAAGTATGAGGCTTTAGAGCAGCCAT 1800
 Qy 1801 TCCATAGGTCCACAGGGCCATCGTGGATGAGTGTCTGATGTTGAGAGAGTCCACT 1860
 Db 1801 TCCATAGGTCCACAGGGCCATCGTGGATGAGTGTCTGATGTTGAGAGAGTCCACT 1860
 Qy 1861 GGGTATTTGGAGGCGCAAGCGCTCCACCGGAGTTAGTGAATGGGTTAGATGAAT 1920
 Db 1861 GGGTATTTGGAGGCGCAAGCGCTCCACCGGAGTTAGTGAATGGGTTAGATGAAT 1920
 Qy 1921 GCGTGGGTCGAAAGCGAGTATGTTTCTGAGAGTGTCCGCACTGTATGCTCTT 1980
 Db 1921 GCGTGGGTCGAAAGCGAGTATGTTTCTGAGAGTGTCCGCACTGTATGCTCTT 1980
 Qy 1981 GCAACGAAGCAAGTCTGACATATTCATCAACCTCAAGGTTCTCTCCCAAGA 2040
 Db 1981 GCAACGAAGCAAGTCTGACATATTCATCAACCTCAAGGTTCTCTCCCAAGA 2040
 Qy 2041 AATTGATATATGCTTTTAAAGTTTGTGATGAAATTTAAAGTTTGTGCTCCGTTAA 2100
 Db 2041 AATTGATATATGCTTTTAAAGTTTGTGATGAAATTTAAAGTTTGTGCTCCGTTAA 2100
 Qy 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAACAGGCTGAAATTCAGTTG 2160
 Db 2101 TGCATCTGTTATGATATATCTATGATTCATTAGGCAAAACAGGCTGAAATTCAGTTG 2160
 Qy 2161 AAATCATACCAAGAGTGTGTAAGAGCTGAGGAGATCTCTGAGGCAATCACAG 2220
 Db 2161 AAATCATACCAAGAGTGTGTAAGAGCTGAGGAGATCTCTGAGGCAATCACAG 2220
 Qy 2221 CTGAACTACTTTAAGAACAGCTCTCAAAAACAGAACAGCAGGCTGCTTGAAGAA 2280
 Db 2221 CTGAACTACTTTAAGAACAGCTCTCAAAAACAGAACAGCAGGCTGCTTGAAGAA 2280
 Qy 2281 TCTCTGGAATTTATGACGAGAGAGCTGCGTAAGAACTGCAAGATTAATCGATCGTGA 2340
 Db 2281 TCTCTGGAATTTATGACGAGAGAGCTGCGTAAGAACTGCAAGATTAATCGATCGTGA 2340
 Qy 2341 CAGAGAACTAAGATGACGAGATGAAACAGAGGAAAGATGATATTTCTTGAAT 2400
 Db 2341 CAGAGAACTAAGATGACGAGATGAAACAGAGGAAAGATGATATTTCTTGAAT 2400
 Qy 2401 CACAACTGACATTTTGTATTAATCTACTGATTCATTTTGTATTAATTTGCAACAA 2460
 Db 2401 CACAACTGACATTTTGTATTAATCTACTGATTCATTTTGTATTAATTTGCAACAA 2460
 Qy 2461 AAAACCTGTGTGTTGAAAGATGATGACACAGCAGGTTTTCATGATTCATCAAA 2520
 Db 2461 AAAACCTGTGTGTTGAAAGATGATGACACAGCAGGTTTTCATGATTCATCAAA 2520
 Qy 2521 CAGATCCATGAAGAAAG 2580
 Db 2521 CAGATCCATGAAGAAAG 2580
 Qy 2581 CGTGCCAAAGTTGTGGCCAGCAGCAGCAGAACATTAATCCCTCTAGCAATGACATTCG 2640
 Db 2581 CGTGCCAAAGTTGTGGCCAGCAGCAGCAGAACATTAATCCCTCTAGCAATGACATTCG 2640
 Qy 2641 CGAAAGAGTATATGTAATCTAATCAATTAATCCCTCTGCGGTTTGTTCAAACCTA 2700
 Db 2641 CGAAAGAGTATATGTAATCTAATCAATTAATCCCTCTGCGGTTTGTTCAAACCTA 2700
 Qy 2701 AGAGTAACGTAATTAATCCGCTTTGATTTCTTTCGAGACCTAGAGAGTCAAGCTTC 2760
 Db 2701 AGAGTAACGTAATTAATCCGCTTTGATTTCTTTCGAGACCTAGAGAGTCAAGCTTC 2760
 Qy 2761 ATGAGTTTCAAGAGAAAGAGATGAGAGATTTGTGAAAGAGAGAGATGATTA 2820
 Db 2761 ATGAGTTTCAAGAGAAAGAGATGAGAGATTTGTGAAAGAGAGAGATGATTA 2820
 Qy 2821 GATCAAGAGAGAGATGAGAGATGAGAGAGAGAGAGATGAGAGATGATTA 2880
 Db 2821 GATCAAGAGAGAGATGAGAGATGAGAGAGAGAGAGATGAGAGATGATTA 2880

Qy 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTAACAGCATGGCTTCAATGAA 2940
 Db 2881 GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTAACAGCATGGCTTCAATGAA 2940
 Qy 2941 GATGATTGAGCAAAAGTCTGTACACAGACAGACATTAAGTTCTTTGTTTCTTTG 3000
 Db 2941 GATGATTGAGCAAAAGTCTGTACACAGACAGACATTAAGTTCTTTGTTTCTTTG 3000
 Qy 3001 GTATGTCGAAAGTGAAGATCTGAGAGACTCCATTTAAATCTAGAGCAAAATCTAAGA 3060
 Db 3001 GTATGTCGAAAGTGAAGATCTGAGAGACTCCATTTAAATCTAGAGCAAAATCTAAGA 3060
 Qy 3061 GATTATGATTTATATCTCCATTTTAAAGAGGATCTAAGAGACATTAAGTTCTT 3120
 Db 3061 GATTATGATTTATATCTCCATTTTAAAGAGGATCTAAGAGACATTAAGTTCTT 3120
 Qy 3121 GTGACTAAAACCAAGTTCTCTGATATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 3180
 Db 3121 GTGACTAAAACCAAGTTCTCTGATATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 3180
 Qy 3181 TTGACATATTTACCAAGCTGACAGATGATCAAGATGCAATGCAATCATGTTT 3240
 Db 3181 TTGACATATTTACCAAGCTGACAGATGATCAAGATGCAATGCAATCATGTTT 3240
 Qy 3241 AGAATTTATATCTACAAATATATGAGTCAAT 3275
 Db 3241 AGAATTTATATCTACAAATATATGAGTCAAT 3275
 RESULT 2
 AAF25374
 ID AAF25374 standard; cDNA; 1878 BP.
 XX
 AC AAF25374;
 DT 15-MAY-2001 (first entry)
 XX
 DE Nucleotide sequence of the Arabidopsis SGS3 polypeptide.
 XX
 KW SGS3 gene; post-transcriptional inactivation; RNA degradation;
 KM viral resistance; resistance; fatty acid content; protein content; ss.
 OS Arabidopsis thaliana.
 XX
 FH Key
 FT CDS 1..1878
 FT /*tag= a
 FT /product= "SGS3"
 XX
 PN W0200105951-A2.
 XX
 PD 25-JAN-2001.
 XX
 PF 13-JUL-2000; 2000MO-FR002052.
 XX
 PR 16-JUL-1999; 99PR-00009417.
 PR 26-JAN-2000; 2000ER-00001006.
 XX
 XX (AVER) AVENTIS CROPS SCIENCE SA.
 PA (INRG) INST NAT RECH AGRONOMIQUE.
 XX
 PI Beclin C, Elmayan T, Vaucheret H;
 XX
 DR WPI; 2001-159529/16.
 DR P-PSDB; AAB31798.
 XX
 PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
 PT resistance in plants and, when inhibited, for increasing transgene
 PT expression.
 XX
 PS Claim 1; Page 32-35; 36pd; French.
 XX

CC The present sequence encodes the Arabidopsis thaliana SGS3 polypeptide.
 CC The SGS3 gene is essential for post-transcriptional inactivation
 CC (degradation of RNA) and for resistance to viruses. Overexpression of
 CC SGS3 results in plants with increased resistance to viruses, while
 CC inactivation of SGS3 in transgenic plants (e.g. by expressing antisense
 CC RNA, by mutation or by homologous recombination) increases the level of
 CC the transgene product. This product may e.g. impart resistance (to
 CC herbicide, insects or pathogens), alter contents of essential fatty acids
 CC or proteins, or is pharmaceutically active, e.g. an immunoglobulin or
 CC interferon
 CC

Sequence 1878 BP; 575 A; 332 C; 554 G; 417 T; 0 U; 0 Other;

Query Match 29.4%; Score 963; DB 4; Length 1878;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 696 ATGAGTTCTAGGGCTGTCTCAATGCTTAAGAAAAAAGCTTCAGGGTGTATAGGCT 755
 DB 1 ATGAGTTCTAGGGCTGTCTCAATGCTTAAGAAAAAAGCTTCAGGGTGTATAGGCT 60
 QY 756 GAGGTTGAACAGTTGGTTCAGAGGAGACGAGACTGGCTTCTTCAAGATGAT 815
 DB 61 GAGGTTGAACAGTTGGTTCAGAGGAGACGAGACTGGCTTCTTCAAGATGAT 120
 QY 816 GAGGAGAGTGGAGGTGCTATTTCCAGAAAGAAACAAACAGAGAAACACTTCTGGA 875
 DB 121 GAGGAGAGTGGAGGTGCTATTTCCAGAAAGAAACAAACAGAGAAACACTTCTGGA 180
 QY 876 AAAAATTGGGTTCTCAGAAATTCGATCTCTAGAGCTTGGGGTGTCTACAGCAAGG 935
 DB 181 AAAAATTGGGTTCTCAGAAATTCGATCTCTAGAGCTTGGGGTGTCTACAGCAAGG 240
 QY 936 AGAGTAGCAACGATCTGGGAGAGAAACAAATGATCCGGAGAGGTAAAGCAATGCT 995
 DB 241 AGAGTAGCAACGATCTGGGAGAGAAACAAATGATCCGGAGAGGTAAAGCAATGCT 300
 QY 996 CGGGGCAATTCAGGCTAAATATCTGCTGGGAGAGAGGCTTGGAGAGAAATGATGATAC 1055
 DB 301 CGGGGCAATTCAGGCTAAATATCTGCTGGGAGAGAGGCTTGGAGAGAAATGATGATAC 360
 QY 1056 AACTTGTGGACCCCACTGTATCTTGCCCTCTTGAAGAGAGATGGAATGGCAG 1115
 DB 361 AACTTGTGGACCCCACTGTATCTTGCCCTCTTGAAGAGAGATGGAATGGCAG 420
 QY 1116 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTCTCTGAGAGTGAAGATGAT 1175
 DB 421 GCAAGAGAGGTTCTGCTCAGACACACAGCTGTGACAGAGTTCTCTGAGAGTGAAGATGAT 480
 QY 1176 GTGATATATGCTTGAAGAGAGATGATCCGATGCTTGAATGATTCATGATGACAC 1235
 DB 481 GTGATATATGCTTGAAGAGAGATGATCCGATGCTTGAATGATTCATGATGACAC 540
 QY 1236 CTTCAGATGATGATTTATGACTCGAGTGAAGTCAAAAGACCATGATCAGAAAGCAG 1295
 DB 541 CTTCAGATGATGATTTATGACTCGAGTGAAGTCAAAAGACCATGATCAGAAAGCAG 600
 QY 1296 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGGATGAGTCTTGTGATGACCAATTAAT 1355
 DB 601 AATAAGTGTCAAAAAGTTCTTTGGCAGCTTGGATGAGTCTTGTGATGACCAATTAAT 660
 QY 1356 GAACACAGAGGAGGAGTGTCTGAGCTTGTGAGAAACGAGACCTGGTCCATGATGAG 1415
 DB 661 GAACACAGAGGAGGAGTGTCTGAGCTTGTGAGAAACGAGACCTGGTCCATGATGAG 720
 QY 1416 TATAAAGTCAACCTCTTACTGACTGAGAGCAAAAGAGAGTGAAGGAGTTAAAGCTC 1475
 DB 721 TATAAAGTCAACCTCTTACTGACTGAGAGCAAAAGAGAGTGAAGGAGTTAAAGCTC 780
 QY 1476 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGAGAGAGGCGCATCTGTATT 1535
 DB 781 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGAGAGAGGCGCATCTGTATT 840

QY 1536 CTTTGTGTGAGATTTATGAGGAGTGAAGGTTTGGGTGAGAGTAAAAAGATTTAGAA 1595
 DB 841 CTTTGTGTGAGATTTATGAGGAGTGAAGGTTTGGGTGAGAGTAAAAAGATTTAGAA 900
 QY 1596 ATTGTCTGAGCTCAATGATCATCATCATGATATAGACTGATTAAGACGATTAAGAT 1655
 DB 901 ATTGTCTGAGCTCAATGATCATCATCATGATATAGACTGATTAAGACGATTAAGAT 960
 QY 1656 AAG 1658
 DB 961 AAG 963

RESULT 3

AAF25372/C

ID AAF25372 standard, DNA, 27 BP.

XX AAF25372;

DT 15-MAY-2001 (first entry)

DE PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.

XX SGS3 gene; post-transcriptional inactivation; RNA degradation;

KW viral resistance; resistance; fatty acid content; protein content;

KW PCR primer; ss.

OS Arabidopsis thaliana.

XX W0200105951-A2.

PN 25-JAN-2001.

PD 13-JUL-2000; 2000MO-FR002052.

PF 16-JUL-1999; 99FR-00009417.

PR 26-JAN-2000; 2000FR-00001006.

XX (AVET) AVENTIS CROSCIENCE SA.

PA (INRG) INST NAT RECH AGRONOMIQUE.

XX Beclin C, Elmayan T, Vaucheret H;

XX WPI; 2001-159529/16.

XX New SGS3 gene from Arabidopsis thaliana, useful for increasing virus

PT resistance in plants and, when inhibited, for increasing transgene

PT expression.

XX Example 1; Page 22; 36pp; French.

XX PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis

CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-

CC transcriptional inactivation (degradation of RNA) and for resistance to

CC viruses. Overexpression of SGS3 results in plants with increased

CC resistance to viruses, while inactivation of SGS3 in transgenic plants

CC (e.g. by expressing antisense RNA, by mutation or by homologous

CC recombination) increases the level of the transgene product. This product

CC may e.g. impart resistance (to herbicide, insects or pathogens), alter

CC contents of essential fatty acids or proteins, or is pharmaceutically

CC active, e.g. an immunoglobulin or interferon

XX Sequence 27 BP; 6 A; 7 C; 5 G; 9 T; 0 U; 0 Other;

Query Match 0.8%; Score 27; DB 4; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.055; Indels 0; Gaps 0;

Matches 27; Conservative 0; Mismatches 0;

QY 2926 GGCCTTCAATGAGATGATTTAGAC 2952

DB 27 GGCCTTCAATGAGATGATTTAGAC 1

```

RESULT 4
AAF25371
ID AAF25371 standard; DNA; 23 BP.
XX
AC AAF25371;
XX
DT 15-MAY-2001 (first entry)
XX
DE PCR primer for cDNA encoding an Arabidopsis SGS3 polypeptide.
XX
KM SGS3 gene; post-transcriptional inactivation; RNA degradation;
KM viral resistance; resistance; fatty acid content; protein content;
KM PCR primer; ss.
XX
OS Arabidopsis thaliana.
XX
PN W0200105951-A2.
XX
PD 25-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-FR002052.
XX
PR 16-JUL-1999; 99FR-00009417.
XX
PR 26-JAN-2000; 2000FR-00001006.
XX
PA (AVET ) AVENTIS CROPS SCIENCE SA.
XX (INRG ) INST NAT RECH AGRONOMIQUE.
XX
PI Beclin C, Elmayan T, Vaucheret H;
XX
DR WPI; 2001-159529/16.
XX
PT New SGS3 gene from Arabidopsis thaliana, useful for increasing virus
PT resistance in plants and, when inhibited, for increasing transgene
PT expression.
XX
PS Example 1; Page 22; 36pp; French.
XX
CC PCR primers AAF25371-72 were used to amplify cDNA encoding an Arabidopsis
CC thaliana SGS3 polypeptide. The SGS3 gene is essential for post-
CC transcriptional inactivation (degradation of RNA) and for resistance to
CC viruses. Overexpression of SGS3 results in plants with increased
CC resistance to viruses, while inactivation of SGS3 in transgenic plants
CC (e.g. by expressing antisense RNA, by mutation or by homologous
CC recombination) increases the level of the transgene product. This product
CC may e.g. impart resistance (to herbicide, insects or pathogens), alter
CC contents of essential fatty acids or proteins, or is pharmaceutically
CC active, e.g. an immunoglobulin or interferon
XX
SQ Sequence 23 BP; 6 A; 4 C; 7 G; 6 T; 0 U; 0 Other;

Query Match 0.7%; Score 23; DB 4; Length 23;
Best Local Similarity 100.0%; Pred. No. 4.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 693 AAATGAGTCTAGGGCTGATCC 715
DB 1 AAATGAGTCTAGGGCTGATCC 23

```

```

XX
OS Homo sapiens.
XX
PN W0200202806-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007470.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIC-) EPIDEMIOLOGY AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154757/20.
XX
PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligonucleotides,
PT useful for detecting cytosine methylation state of genes associated with
PT pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS Claim 1; SEQ ID NO 123; 24pp; English.
XX
CC The invention relates to a nucleic acid comprising a sequence at least 18
CC bases in length of a segment of the chemically pretreated DNA of genes
CC associated with pharmacogenomics according to one of the sequences of the
CC genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996),
CC NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899 and
CC their complementary sequences, or a sequence (S1) chosen from 87
CC sequences and their complements. The chemical pretreatment is bisulphite
CC treatment to convert cytosines (but not methyl-cytosines) into uracils.
CC Also included are an oligomer (II) in particular an oligonucleotide or a
CC peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC base sequence having a length of 9 nucleotides which hybridises to or is
CC identical to a chemically pretreated DNA of genes associated with
CC pharmacogenomics and their complements, arranged in an array for
CC analysing diseases associated with the methylation state (CpG) and/or
CC detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC and their complements is useful for diagnosis and therapy of solid
CC tumours and cancer. The present sequence represents one the 87 DNA
CC sequences or its complement. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6136 BP; 1661 A; 69 C; 1415 G; 2991 T; 0 U; 0 Other;

Query Match 0.7%; Score 22; DB 6; Length 6136;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3145 TATTGCTTTTGTGTTAA 3166
DB 2122 TATTGCTTTTGTGTTAA 2143

```

```

RESULT 5
ABK40041
ID ABK40041 standard; DNA; 6136 BP.
XX
AC ABK40041;
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #62 strand 1.
XX
KM Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KM cytosine; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KM UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.

```

```

RESULT 6
ABN80100
ID ABN80100 standard; DNA; 6301 BP.
XX
AC ABN80100;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 117.
XX
KM Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KM heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KM dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;

```

KM antidiabetic; cytostatic; anticonvulsant; ds.
 XX Homo sapiens.
 OS Synthetic.
 XX WO200200927-A2.
 XX 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP007536.
 XX
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPiG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130908/17.
 XX
 XX Novel nucleic acid useful for diagnosis and therapy of diseases
 PT associated with development genes such as diabetes, comprises a sequence
 PT of a segment of chemically pretreated DNA of genes associated with
 PT development.
 XX
 XX Claim 1; SEQ ID NO 117; 27pp; English.
 XX
 CC The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)
 CC of genes associated with development selected from 87 genes listed in the
 CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
 CC sequences (ABN79984-ABN80333) or their complements. The invention is
 CC useful for the diagnosis or therapy of diseases associated with
 CC development genes, in particular disease related to homeobox containing
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
 CC associated with congenital heart disease, epilepsy, diseases related to
 CC histone deacetylation, Currairino syndrome, diseases related with the
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.
 CC Oligomers specific to each of the genes are useful for detecting the
 CC methylation state of all CpG dinucleotides within the 350 sequences or
 CC (II) and their complementary sequences, as primer oligonucleotides for
 CC the amplification of the 350 sequences, (II) and/or their complements and
 CC as oligomer probes for detecting the cytosine methylation state and/or
 CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
 CC patent did not form part of the printed specification but is based on
 CC sequence information supplied to Derwent by the European Patent Office
 XX
 SQ Sequence 6301 BP; 1787 A; 59 C; 1247 G; 3207 T; 0 U; 1 Other;
 Query Match 0.7%; Score 22; DB 6; Length 6301;
 Best Local Similarity 100.0%; Pred. No. 12;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3142 TAGTATTTGTTTTTTTGGT 3163
 DB 3729 TAGTATTTGTTTTTTTGGT 3750
 RESULT 7
 ABL33241
 ID ABL33241 standard; DNA; 9888 BP.
 XX
 XX ABL33241;
 AC
 XX 26-MAR-2002 (first entry)
 XX
 XX Human immune system associated gene SEQ ID NO: 1214.
 DE
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KM antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KM

KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 ds.
 XX Homo sapiens.
 OS
 XX WO200200928-A2.
 XX 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP007537.
 XX
 XX 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 XX (EPiG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful for
 PT diagnosis and treatment of diseases associated with abnormal cytosine
 PT methylation.
 PT
 XX
 XX Claim 1; SEQ ID NO 1214; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention
 XX
 SQ Sequence 9888 BP; 2830 A; 160 C; 2148 G; 4750 T; 0 U; 0 Other;
 Query Match 0.7%; Score 22; DB 6; Length 9888;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3142 TAGTATTTGTTTTTTTGGT 3163
 DB 8219 TAGTATTTGTTTTTTTGGT 8240
 RESULT 8
 ABL32509
 ID ABL32509 standard; DNA; 6048 BP.
 XX
 XX ABL32509;
 AC
 XX 26-MAR-2002 (first entry)
 XX
 XX Human immune system associated gene SEQ ID NO: 482.
 DE
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KM antiarteriosclerotic; antianaemic; cytostatic; nootropic;
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KM antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
 KM ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200200928-A2.
 XX 03-JAN-2002.
 XX
 XX 02-JUL-2001; 2001WO-EP007537.
 XX


```

XX 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX diagnosis and treatment of diseases associated with abnormal cytosine
XX methylation.
XX
XX Claim 1; SEQ ID NO 482; 32bp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 6048 BP; 1296 A; 103 C; 1361 G; 3288 T; 0 U; 0 Other;
XX
XX Query Match 0.6%; Score 21; DB 6; Length 6048;
XX Best Local Similarity 100.0%; Pred. No. 34;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 526 TTGGTTTGGTTTGGTTT 546
XX |||||
XX 5395 TTGGTTTGGTTTGGTTT 5415
XX
XX RESULT 9
XX ABL32031
XX ID ABL32031 standard; DNA; 6072 BP.
XX
XX ABL32031;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 4.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; anti-anaemic; cytosatic; noctropic;
XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
XX ds.
XX
XX Homo sapiens.
XX
XX WO200200928-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP007537.
XX
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2002-130909/17.
XX
XX Nucleic acid comprising fragment of chemically modified gene, useful for
XX

```

```

PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
XX Claim 1; SEQ ID NO 4; 32bp + Sequence Listing; German.
XX
XX The present invention provides a number of human immune system associated
XX genes which are modified by the methylation of cytosines. The sequences
XX can be used in the diagnosis and treatment of immune system disorders,
XX including eye diseases such as retinopathy, neovascular glaucoma and
XX macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
XX leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
XX rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
XX diseases. The present sequence is a gene of the invention
XX
XX Sequence 6072 BP; 1595 A; 155 C; 1384 G; 2938 T; 0 U; 0 Other;
XX
XX Query Match 0.6%; Score 21; DB 6; Length 6072;
XX Best Local Similarity 100.0%; Pred. No. 34;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 3145 TATTGTTTGGTTTGGTTAA 3165
XX |||||
XX 3717 TATTGTTTGGTTTGGTTAA 3737
XX
XX RESULT 10
XX AAS46448
XX ID AAS46448 standard; DNA; 8245 BP.
XX
XX AAS46448;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #170.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytosatic; cancer;
XX tumour; CPG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP002955.
XX
XX 15-MAR-2000; 2000DE-01013847.
XX 06-APR-2000; 2000DE-01019058.
XX 07-APR-2000; 2000DE-01019173.
XX 30-JUN-2000; 2000DE-01032529.
XX 01-SEP-2000; 2000DE-01043826.
XX
XX (EPiG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumor suppressor
XX genes and oncogenes, useful in designing primers and probes for analyzing
XX diseases associated with cytosine methylation state e.g. cancer.
XX
XX Claim 1; SEQ ID NO 170; 27bp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and oncogenes
XX having a sequence taken from 536 (actually 533 since numbers 408, 458 and
XX 500 are missing from the sequence listing) sequences (58) and sequences
XX complementary to (58). The nucleic acid may be a peptide nucleic acid-
XX oligomer (PNA) of at least 9 nucleotides and may form part of a set of
XX probes for detecting the cytosine methylation state and/or single
XX

```

CC nucleotide polymorphisms and also to be used in an array for analysing
 CC diseases associated with Cpg dinucleotides e.g. cancers and tumours. The
 CC probes can also be used in a method for ascertaining genetic and/or
 CC epigenetic parameters for the diagnosis and/or therapy of existing
 CC diseases or the predisposition to specific diseases, by analysing
 CC cytosine methylations. The parameters may be compared to another set of
 CC genetic and/or epigenetic parameters, the differences serving as basis
 CC for diagnosis and/or prognosis events which are disadvantageous to
 CC patients. The present sequence is one of the 533 genomic sequences
 CC derived from tumour suppressor genes and oncogenes. Sequences with even
 CC numbered Seq ID numbers are the complementary sequence of the
 CC corresponding odd numbered sequence (e.g. ID 2 and ID1, ID 536 and ID
 CC 535, except for those whose partner sequence is missing). Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences

SO Sequence 8245 BP; 1978 A; 305 C; 2119 G; 3842 T; 0 U; 1 Other;

Query Match 0.6%; Score 21; DB 4; Length 8245;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3141 TTAGTATTTTGTGTTTTTTTG 3161
 |||||
 DB 6632 TTAGTATTTTGTGTTTTTTTG 6652

RESULT 11
 ABN80176
 ID ABN80176 standard; DNA; 13453 BP.
 XX ABN80176;
 AC
 XX
 XX
 DT 15-JUL-2002 (first entry)
 DE Human chemically modified disease associated gene SEQ ID NO 193.
 XX
 XX Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
 KM heart disease; epilepsy; histone deacetylation; muscular dystrophy;
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
 KM antidiabetic; cytostatic; anticonvulsant; ds.
 XX
 XX Homo sapiens.
 OS Synthetic.
 OS
 XX
 PN WO200200927-A2.
 XX
 PD 03-JAN-2002.
 XX
 XX
 PF 02-JUL-2001; 2001WO-EP007536.
 XX
 PR 30-JUN-2000; 2000DE-01032529.
 PR 01-SEP-2000; 2000DE-01043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130908/17.
 XX
 PT Novel nucleic acid useful for diagnosis and therapy of diseases
 PT associated with development genes such as diabetes, comprises a sequence
 PT of a segment of chemically pretreated DNA of genes associated with
 PT development.
 PT
 XX
 XX Claim 1; SEQ ID NO 193; 27pp; English.
 CC
 CC The invention relates to a nucleic acid (I) comprising a sequence at
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)
 CC of genes associated with development selected from 87 genes listed in the
 CC specification such as ACCFN, ADFN, or AFDI and comprising one of 350
 CC sequences (ABN79984-ABN80333) or their complements. The invention is

CC useful for the diagnosis or therapy of diseases associated with
 CC development genes, in particular diseases related to homeobox containing
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
 CC associated with congenital heart disease, epilepsy, diseases related to
 CC histone deacetylation, Currying syndrome, diseases related with the
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.
 CC Oligomers specific to each of the genes are useful for detecting the
 CC methylation state of all Cpg dinucleotides within the 350 sequences or
 CC (II) and their complementary sequences, as primer oligonucleotides for
 CC the amplification of the 350 sequences, (II) and/or their complements and
 CC as oligomer probes for detecting the cytosine methylation state and/or
 CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
 CC patent did not form part of the printed specification but is based on
 CC sequence information supplied to Derwent by the European Patent Office

SO Sequence 13453 BP; 2569 A; 384 C; 3712 G; 6788 T; 0 U; 0 Other;

Query Match 0.6%; Score 21; DB 6; Length 13453;
 Best Local Similarity 100.0%; Pred. No. 34;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3141 TTAGTATTTTGTGTTTTTTTG 3161
 |||||
 DB 7909 TTAGTATTTTGTGTTTTTTTG 7929

RESULT 12
 ADA02738/c
 ID ADA02738 standard; DNA; 52302 BP.
 XX ADA02738;
 AC
 XX
 XX
 DT 06-NOV-2003 (first entry)
 DE Human CCND2 carcinoma associated gene, SEQ ID NO:1256.
 XX
 XX Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
 KM prostate; lymphoma; leukemia; cytostatic; gene therapy; drug screening;
 KM gene; ds.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO2003057146-A2.
 XX
 PD 17-JUL-2003.
 XX
 XX
 PF 26-DEC-2002; 2002WO-US041414.
 XX
 PR 26-DEC-2001; 2001US-00035832.
 XX
 PA (SAGR-) SAGRES DISCOVERY.
 XX
 PI Morris DW;
 XX
 DR WPI; 2003-587068/55.
 XX
 PT New recombinant nucleic acid encoding carcinoma associated protein,
 PT useful for preparing compositions for treating carcinomas.
 PT
 XX
 XX Claim 1; SEQ ID NO 1256; 245pp; English.
 CC
 CC The invention relates to recombinant carcinoma associated (CA) nucleic
 CC acid sequences from mouse and human (ADA01482-ADA03094), and to
 CC recombinant carcinoma associated proteins (CAP) encoded by them. The
 CC invention also encompasses expression vectors and host cells comprising a
 CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
 CC binds to the protein, and a biochip comprising CA nucleic acid or
 CC fragments thereof. The sequences of the invention were identified using
 CC oncogenic retroviruses, which insert into the genome of the host organism
 CC at random. Many of these do not carry transduced host oncogenes or
 CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
 CC direct consequence of the effects of proviral integration into host
 CC protooncogenes. The CA nucleic acid sequences can be used to diagnose

CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 8; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2981 GTTCTTTGTTTGGCTTTGG 3001
11586 GTTCTTTGTTTGGCTTTGG 11566

RESULT 13

AD872476/c
ID ADB72476 standard; DNA; 52302 BP.

XX ADB72476;

XX 04-DEC-2003 (first entry)

XX Human CCND2 gene.

XX human; ds: cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
XX cancer; neoplasm; adenocarcinoma; sarcoma; gene.

XX Homo sapiens.

XX MO2003008583-A2.

XX 30-JAN-2003.

XX 26-DEC-2001; 2001MO-US051291.

XX 02-MAR-2001; 2001US-00798586.

XX 23-OCT-2001; 2001US-00004113.

XX 08-NOV-2001; 2001US-00052482.

XX 30-NOV-2001; 2001US-00997722.

XX 20-DEC-2001; 2001US-00034650.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-239337/23.

XX New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
XX PT cancers, neoplasm, adenocarcinoma, or sarcomas.

XX Claim 1; SEQ ID NO 304; 2304BP; English.

XX The invention relates to a novel recombinant nucleic acid comprising a
XX CC nucleotide sequence selected from any of the 660 sequences fully defined
XX CC in the specification. A polynucleotide of the invention has cytostatic
XX CC activity, and may have a use in gene therapy, or in a vaccine. The
XX CC recombinant nucleic acids and polypeptides are useful for treating
XX CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
XX CC sarcomas. The present sequence represents a human gene of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2981 GTTCTTTGTTTGGCTTTGG 3001
11586 GTTCTTTGTTTGGCTTTGG 11566

RESULT 14

AD85218/c
ID ADC85218 standard; DNA; 52302 BP.

XX ADC85218;

XX 01-JAN-2004 (first entry)

XX Human CCnd2 genomic sequence.

XX Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
XX secreted; transmembrane; intracellular; ds.

XX Homo sapiens.

XX MO2003045230-A2.

XX 05-JUN-2003.

XX 02-DEC-2002; 2002MO-US038582.

XX 30-NOV-2001; 2001US-00997722.

XX (SAGR-) SAGRES DISCOVERY.

XX PI Morris DW, Engelhard EK;

XX WPI; 2003-513603/48.

XX New recombinant nucleic acid comprising a nucleotide sequence of any of
XX PT the carcinoma-associated (CA) genes, useful for screening for drug
XX PT candidates for diagnosing or treating carcinomas.

XX Claim 1; SEQ ID NO 4; 983BP; English.

XX The invention relates to a recombinant nucleic acid comprising a
XX CC nucleotide sequence selected from any of the fully defined carcinoma-
XX CC associated (CA) genes from the 50 tables given in the specification. The
XX CC CA proteins are secreted, transmembrane or intracellular proteins. The
XX CC recombinant nucleic acids are useful for screening for drug candidates
XX CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
XX CC ADC85514 represent CA genes of the invention.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2981 GTTCTTTGTTTGGCTTTGG 3001
11586 GTTCTTTGTTTGGCTTTGG 11566

RESULT 15

AAC52596/c
ID AAC52596 standard; DNA; 420 BP.

XX AAC52596;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana DNA fragment SEQ ID NO: 71849.

XX Hybridisation assay; genetic mapping; gene expression control;
XX KW protein identification; signal transduction pathway; metabolic pathway;
XX KW promoter; termination sequence; ss.

XX Sequence 52302 BP; 13145 A; 12501 C; 12728 G; 13888 T; 0 U; 40 Other;

Query Match 0.6%; Score 21; DB 9; Length 52302;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2981 GTTCTTTGTTTGGCTTTGG 3001
11586 GTTCTTTGTTTGGCTTTGG 11566

```
XX Arabidopsis thaliana.
OS
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-00301439.
XX
PR 25-FEB-1999; 99US-0121825P.
PR 05-MAR-1999; 99US-0121180P.
PR 09-MAR-1999; 99US-0123548P.
PR 23-MAR-1999; 99US-0125788P.
PR 25-MAR-1999; 99US-0126264P.
PR 29-MAR-1999; 99US-0126785P.
PR 01-APR-1999; 99US-0127462P.
PR 06-APR-1999; 99US-0128234P.
PR 08-APR-1999; 99US-0128714P.
PR 16-APR-1999; 99US-0129845P.
PR 19-APR-1999; 99US-0130077P.
PR 21-APR-1999; 99US-0130510P.
PR 23-APR-1999; 99US-0130891P.
PR 28-APR-1999; 99US-0131449P.
PR 30-APR-1999; 99US-0132048P.
PR 04-MAY-1999; 99US-0132407P.
PR 05-MAY-1999; 99US-0132484P.
PR 06-MAY-1999; 99US-0132485P.
PR 06-MAY-1999; 99US-0132486P.
PR 07-MAY-1999; 99US-0132487P.
PR 11-MAY-1999; 99US-0132863P.
PR 14-MAY-1999; 99US-0134256P.
PR 14-MAY-1999; 99US-0134218P.
PR 14-MAY-1999; 99US-0134219P.
PR 14-MAY-1999; 99US-0134221P.
PR 14-MAY-1999; 99US-0134370P.
PR 18-MAY-1999; 99US-0134768P.
PR 19-MAY-1999; 99US-0134941P.
PR 20-MAY-1999; 99US-0135124P.
PR 21-MAY-1999; 99US-0135353P.
PR 24-MAY-1999; 99US-0135629P.
PR 25-MAY-1999; 99US-0136021P.
PR 27-MAY-1999; 99US-0136392P.
PR 28-MAY-1999; 99US-0136782P.
PR 01-JUN-1999; 99US-0137222P.
PR 03-JUN-1999; 99US-0137528P.
PR 04-JUN-1999; 99US-0137502P.
PR 07-JUN-1999; 99US-0137724P.
PR 08-JUN-1999; 99US-0138094P.
PR 10-JUN-1999; 99US-0138540P.
PR 14-JUN-1999; 99US-0138847P.
PR 16-JUN-1999; 99US-0139119P.
PR 16-JUN-1999; 99US-0139452P.
PR 17-JUN-1999; 99US-0139453P.
PR 18-JUN-1999; 99US-0139454P.
PR 18-JUN-1999; 99US-0139455P.
PR 18-JUN-1999; 99US-0139456P.
PR 18-JUN-1999; 99US-0139457P.
PR 18-JUN-1999; 99US-0139458P.
PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 24-JUN-1999; 99US-0140354P.
PR 26-JUN-1999; 99US-0140695P.
PR 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 01-JUL-1999; 99US-0142154P.
PR 02-JUL-1999; 99US-0142055P.
PR 06-JUL-1999; 99US-0142390P.
PR 08-JUL-1999; 99US-0142803P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143624P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144684P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 22-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145913P.
PR 27-JUL-1999; 99US-0145918P.
PR 28-JUL-1999; 99US-0145919P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 17-AUG-1999; 99US-0149368P.
PR 18-AUG-1999; 99US-0149175P.
PR 20-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.
PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151068P.
PR 27-AUG-1999; 99US-0151066P.
PR 30-AUG-1999; 99US-0151080P.
PR 31-AUG-1999; 99US-0151303P.
PR 01-SEP-1999; 99US-0151438P.
PR 07-SEP-1999; 99US-0151930P.
PR 10-SEP-1999; 99US-0152363P.
PR 99US-0153070P.
```

PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155659P.
PR 28-SEP-1999; 99US-0156458P.
PR 29-SEP-1999; 99US-0156596P.
PR 04-OCT-1999; 99US-0157117P.
PR 05-OCT-1999; 99US-0157753P.
PR 06-OCT-1999; 99US-0157865P.
PR 07-OCT-1999; 99US-0158029P.
PR 08-OCT-1999; 99US-0158232P.
PR 12-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 13-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159329P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 22-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 26-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query March 0.6%; Score 20; DB 3; Length 420;
Best Local Similarity 100.0%; Pred.No.1.le+02;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ACAACAAACAAATTAG 21
|||
Db 378 ACAACAAACAAATTAG 359

Search completed: March 30, 2004, 09:26:44
Job time : 1138.92 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2004, 03:15:12 ; Search time 12253.4 Seconds

(without alignments)
11584.370 Million cell updates/sec

Title: US-10-030-829-1

Perfect score: 3275

Sequence: 1 gcaacaacaacaataa.....caaatatattggtacacaaat 3275

Scoring table: OLIGO_NUC

Gapop 60.0, Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size: 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

```

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_pi:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rtd:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3275	100.0	3275	6	AX078760	AX078760 Sequence
2	3275	100.0	81365	8	AB025633	AB025633 Arabidops
3	2254	68.8	2254	8	AF239719	AF239719 Arabidops
4	972	29.7	2162	8	BT002944	BT002944 Arabidops
5	963	29.4	1878	6	AX078761	AX078761 Sequence
6	963	29.4	1909	8	BT004380	BT004380 Arabidops
7	541	16.5	650	8	ATH528171	ATH528171 Arabidops
8	445	14.4	45	8	ATH527741	ATH527741 Arabidops
9	24	0.7	170765	2	AC102284	AC102284 Mus muscu
10	23	0.7	554	3	AF057056	AF057056 Uroleucon
11	23	0.7	101080	2	BX004860	BX004860 Dario rer
12	23	0.7	101241	5	AL845510	AL845510 Zebrafish
13	23	0.7	126716	9	AC107028	AC107028 Homo sapi
14	23	0.7	152345	5	BX005128	BX005128 Zebrafish
15	23	0.7	180925	2	AC048381	AC048381 Homo sapi
16	22	0.7	2100	3	D84313	D84313 Drosophila
17	22	0.7	6136	6	AX348428	AX348428 Sequence
18	22	0.7	6301	6	AX344692	AX344692 Sequence
19	22	0.7	9888	6	AX346143	AX346143 Sequence
20	22	0.7	80318	5	AL645798	AL645798 Zebrafish
21	22	0.7	81835	8	ATFC9	ATFC9 Arabidops
22	22	0.7	110000	2	AC107170_1	297344 Arabidops
23	22	0.7	133120	9	HS333P23	Continuation (2 of
24	22	0.7	141509	9	AC006044	295400 Human DNA s
25	22	0.7	150030	2	AC145578	AC006044 Homo sapi
26	22	0.7	154923	9	AC026185	AC026185 Mus muscu
27	22	0.7	158661	10	AC074229	AC074229 Mus sapi
28	22	0.7	174612	9	AC018494	AC018494 Homo sapi
29	22	0.7	174641	2	AC016346	AC016346 Homo sapi
30	22	0.7	175569	22	AC109214	AC109214 Mus muscu
31	22	0.7	177529	2	AC119271	AC119271 Mus muscu
32	22	0.7	177834	9	AC018811	AC018811 Homo sapi
33	22	0.7	178928	2	AC121725	AC121725 Rattus no
34	22	0.7	179134	2	AC119607	AC119607 Rattus no
35	22	0.7	181166	9	AC087897	AC087897 Homo sapi
36	22	0.7	182509	9	AC112211	AC112211 Homo sapi
37	22	0.7	183936	10	AC007978	AC007978 Mus muscu
38	22	0.7	191391	2	AC128938	AC128938 Rattus no
39	22	0.7	192173	9	AC107462	AC107462 Homo sapi
40	22	0.7	193523	5	BX001014	BX001014 Zebrafish
41	22	0.7	197909	10	AL928719	AL928719 Mouse DNA
42	22	0.7	198067	8	ATCHRIV47	AL161547 Arabidops
43	22	0.7	201208	2	AC107451	AC107451 Mus muscu
44	22	0.7	201803	2	AC113572	AC113572 Canis fam
45	22	0.7	205209	2	AC137942	AC137942 Mus muscu

ALIGNMENTS

```

RESULT 1
AX078760 3275 bp DNA linear PAT 22-FEB-2001
LOCUS AX078760
DEFINITION Sequence 1 from Patent WO0105951.
ACCESSION AX078760
VERSION AK078760.1 GI:13158379
KEYWORDS
SOURCE
ORGANISM
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 Becilj,C., Elmayer,T. and Vaucheret,H.
AUTHORS Novel gsg3 plant gene and use thereof
TITLE

```

JOURNAL Patent: WO 0105951-A 1 25-JAN-2001;
AVENTIS CROPS SCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
AGRONOMIQUE (FR)
FEATURES Location/Qualifiers
source 1..3275
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
primer_bind 693..715
/note="P356AD"
primer_bind complement(2926..2952)
/note="P356Y"
ORIGIN
Query Match 100.0%; Score 3275; DB 6; Length 3275;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 GACAAACAAACAAATTAAGCAAGTCATGTCGATGCAATTAATTAAGTGGAAACA 60
Qy 1 TTAAGTAAAGCAAAAGAAAAAGGTAACAAAATGAAAACAAATCAATCTGAT 120
Db 61 TTAAGTAAAGCAAAAGAAAAAGGTAACAAAATGAAAACAAATCAATCTGAT 120
Qy 121 GAAATTTGGAGTCCAGATGCGAAAAAGAGCCGTTTGAAGCTTAATTAAGCTTCTC 180
Db 121 GAAATTTGGAGTCCAGATGCGAAAAAGAGCCGTTTGAAGCTTAATTAAGCTTCTC 180
Qy 181 ATTGTCTCTCTTCGTGATTTATTTCTCTCCGAGAGCTGACTCACTACTCTAC 240
Db 181 ATTGTCTCTTCGTGATTTATTTCTCTCCGAGAGCTGACTCACTACTCTAC 240
Qy 241 TCTCCGCGCTTTAACTTACGTTCTCCGCTGTTACTCTGTAAGTTTCTGCTTAAG 300
Db 241 TCTCCGCGCTTTAACTTACGTTCTCCGCTGTTACTCTGTAAGTTTCTGCTTAAG 300
Qy 301 CCTCGATCGCTCACCGCATGCAATTCGTGCTGATTTCTCTTTCTTCGCTGAAAA 360
Db 301 CCTCGATCGCTCACCGCATGCAATTCGTGCTGATTTCTCTTTCTTCGCTGAAAA 360
Qy 361 ATTGCCCTAATGTTCTGATTTGCAAGTTTGTGCTATGAGTTTCTCTTAT 420
Db 361 ATTGCCCTAATGTTCTGATTTGCAAGTTTGTGCTATGAGTTTCTCTTAT 420
Qy 421 TTTTAAAGTTTGAAGTACGATACCTGCGCTTACTGTTTGTGTTTGTGTTGCT 480
Db 421 TTTTAAAGTTTGAAGTACGATACCTGCGCTTACTGTTTGTGTTTGTGTTGCT 480
Qy 481 TTCAACGTTTGAAGTACGATACCTGCGCTTACTGTTTGTGTTTGTGTTGCT 540
Db 481 TTCAACGTTTGAAGTACGATACCTGCGCTTACTGTTTGTGTTTGTGTTGCT 540
Qy 541 TTGTTCAATTAATTCGATTCATCTTTTGTGTTGATTTGTTTGTGAGCT 600
Db 541 TTGTTCAATTAATTCGATTCATCTTTTGTGTTGATTTGTTTGTGAGCT 600
Qy 601 ATGGCTTGGCTGTTTGAATTAATTCGATTCATCTTTTGTGTTGATTTGTTGAGCT 660
Db 601 ATGGCTTGGCTGTTTGAATTAATTCGATTCATCTTTTGTGTTGATTTGTTGAGCT 660
Qy 661 ACTGTGGCTTCTTGTGCTATAGTTGTAATAATGAGTTCTTGAAGGCTGTCGAATG 720
Db 661 ACTGTGGCTTCTTGTGCTATAGTTGTAATAATGAGTTCTTGAAGGCTGTCGAATG 720
Qy 721 CTAAAGAAAAGACGTTCAAGGCTGTTAAGAGCTGAGTTGAACAGTTGTTCAAGTT 780
Db 721 CTAAAGAAAAGACGTTCAAGGCTGTTAAGAGCTGAGTTGAACAGTTGTTCAAGTT 780
Qy 781 TGGAGGAGGAGAGCTGGCTTCTTCAAGATGATGAGAGAGTGGAGAGTCAATTTCCA 840
Db 781 TGGAGGAGGAGAGCTGGCTTCTTCAAGATGATGAGAGAGTGGAGAGTCAATTTCCA 840

Qy 841 AGAAGAAACAAGAAACAACGAGAAACAATTTCTGAAAAAATTTGGTTTCTCAATTTCA 900
Db 841 AGAAGAAACAAGAAACAACGAGAAACAATTTCTGAAAAAATTTGGTTTCTCAATTTCA 900
Qy 901 ATCTCTCTAAGCTTGGGGTGTGTCACACGAAAGGAGAGTATGCAATCTTGGAGAG 960
Db 901 ATCTCTCTAAGCTTGGGGTGTGTCACACGAAAGGAGAGTATGCAATCTTGGAGAG 960
Qy 961 GAAACATGATATCCGGAGAGAGTAAAGCAATGTTGGGGCATTCAAGCTTAATCATATCTG 1020
Db 961 GAAACATGATATCCGGAGAGAGTAAAGCAATGTTGGGGCATTCAAGCTTAATCATATCTG 1020
Qy 1021 GTGGGGAGAGAGCTTGAAGCAAGAAATGATTAACAATTTGTGCAACCCCACTGTAT 1080
Db 1021 GTGGGGAGAGAGCTTGAAGCAAGAAATGATTAACAATTTGTGCAACCCCACTGTAT 1080
Qy 1081 CTGCGCTCTCTTGAAGAGAGTGAATTTGGAGGCAAGAGAGTTCGCTCAGCACA 1140
Db 1081 CTGCGCTCTCTTGAAGAGAGTGAATTTGGAGGCAAGAGAGTTCGCTCAGCACA 1140
Qy 1141 CAGCTGTGAGAGATTCTGACGTGAGAGATGATGATTAATGCTTCTGAGAGAGAGA 1200
Db 1141 CAGCTGTGAGAGATTCTGACGTGAGAGATGATGATTAATGCTTCTGAGAGAGAGA 1200
Qy 1201 ATGATTCGATGCTTTGATGATTTCTGATGACGACTTCAAGTATGATGATGATGATG 1260
Db 1201 ATGATTCGATGCTTTGATGATTTCTGATGACGACTTCAAGTATGATGATGATGATG 1260
Qy 1261 ATGTAGTCAAAAGAGCAATGATCAAGAAAGCAATTAAGTGTCAAAAAGTCTTGG 1320
Db 1261 ATGTAGTCAAAAGAGCAATGATCAAGAAAGCAATTAAGTGTCAAAAAGTCTTGG 1320
Qy 1321 GCACTGTGATGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
Db 1321 GCACTGTGATGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1380
Qy 1381 CAGCTGTGAGAGAGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 1440
Db 1381 CAGCTGTGAGAGAGAGCTTGTGATGATGATGATGATGATGATGATGATGATGATG 1440
Qy 1441 ATGCGAGACAAAGAGAGCTAGAGAGTAACTCAATGAGAAATTTGGCTGAAGTTTAA 1500
Db 1441 ATGCGAGACAAAGAGAGCTAGAGAGTAACTCAATGAGAAATTTGGCTGAAGTTTAA 1500
Qy 1501 AAAAGATCTAAGATGAGAGAGAGCTGATGATGATGATGATGATGATGATGATGATG 1560
Db 1501 AAAAGATCTAAGATGAGAGAGAGCTGATGATGATGATGATGATGATGATGATGATG 1560
Qy 1561 GGAAGGCTTGGGTGAGAGATGAAAGATTAAGAAATGCTCGGCTCCAAATGTCATCA 1620
Db 1561 GGAAGGCTTGGGTGAGAGATGAAAGATTAAGAAATGCTCGGCTCCAAATGTCATCA 1620
Qy 1621 TCATGAATCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
Db 1621 TCATGAATCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
Qy 1681 TTTCTTAATTTTCTTCTGATTTCTGATGATGATGATGATGATGATGATGATGATGATG 1740
Db 1681 TTTCTTAATTTTCTTCTGATTTCTGATGATGATGATGATGATGATGATGATGATGATG 1740
Qy 1741 ATGGGCAACAAAGAGAGTGTGATTAATTCGACAGATTAAGGCTTAAAGAGAGAGCAT 1800
Db 1741 ATGGGCAACAAAGAGAGTGTGATTAATTCGACAGATTAAGGCTTAAAGAGAGAGCAT 1800
Qy 1801 TCTATATGTCACAGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
Db 1801 TCTATATGTCACAGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1860
Qy 1861 GGCATATTTGAGAGCGGAAGGCTTCAACGGAGGTTAGCTGATGATGATGATGATGATG 1920
Db 1861 GGCATATTTGAGAGCGGAAGGCTTCAACGGAGGTTAGCTGATGATGATGATGATGATG 1920

QY	1921	GCCTGGGGGCACAAGCGAGATATGTTTTCTGGAGGGTGTTCGCACTGTATAGGCTTCTT	1980
Db	1921	GCTTGGGGGCAGAAACGGAGTATGTTTTCTGGAGGGTGTTCGCACTGTATAGGCTTCTT	1980
QY	1981	GCAAGGAAGCAGAGATCTGGACATATTCATCAACCTCTCAAGGTTCTCTCCCCAAGA	2048
Db	1981	GCAAGGAAGCAGAGATCTGGACATATTCATCAACCTCTCAAGGTTCTCTCCCCAAGA	2048
QY	2041	AATTTGATATATGCTTTTATGTTTTGTCAATGGATTTTAAAGTTTGTGTCCGTATTA	2100
Db	2041	AATTTGATATATGCTTTTATGTTTTGTCAATGGATTTTAAAGTTTGTGTCCGTATTA	2100
QY	2101	TGCATCTGTATATGTTATATATCTAATCATCTAATAGCAAAACAAGCTAAATTCGACTTG	2160
Db	2101	TGCATCTGTATATGTTATATATCTAATCATCTAATAGCAAAACAAGCTAAATTCGACTTG	2160
QY	2161	AAATCATATCCAGAGATGGTTGTATAAGAGCTGAGGCGAGATCTCTGAGGAACAATCAGAG	2220
Db	2161	AAATCATATCCAGAGATGGTTGTATAAGAGCTGAGGCGAGATCTCTGAGGAACAATCAGAG	2220
QY	2221	CTGAATCTATCTTTAAGACAAGCTCTCAAAAAGAAACAAGCAGCCCAAGTGTTCGAGAA	2280
Db	2221	CTGAATCTATCTTTAAGACAAGCTCTCAAAAAGAAACAAGCAGCCCAAGTGTTCGAGAA	2280
QY	2281	TCTCTGGAAATTTATAGGAGCAAGCTGGGTAGAACTCGAGAGATATATCCGATGTGTGA	2340
Db	2281	TCTCTGGAAATTTATAGGAGCAAGCTGGGTAGAACTCGAGAGATATATCCGATGTGTGA	2340
QY	2341	CAGAGAACTAAGATCAGATCAGTAAACAGAACAGGGAAGGTATGATTTTCTCTGAAGAT	2400
Db	2341	CAGAGAACTAAGATCAGATCAGTAAACAGAACAGGGAAGGTATGATTTTCTCTGAAGAT	2400
QY	2401	CACAAACTTGAACATTTTGTATTTCCCTACTGTATTCATCATTTTGTATATATGTCACAA	2460
Db	2401	CACAAACTTGAACATTTTGTATTTACTCTATGATTCACATTTTGTATATATGTCACAA	2460
QY	2461	AAAACTGTGGTGTTTGAAGTGGATGACAACAAGCTTTTTCATGTGATCAATCAAA	2520
Db	2461	AAAACTGTGGTGTTTGAAGTGGATGACAACAAGCTTTTTCATGTGATCAATCAAA	2520
QY	2521	CAGATCCATGAAGAAGAGACGCAAAAGAGAGAAATTCGAGATGTTCACACAGAGAA	2580
Db	2521	CAGATCCATGAAGAAGAGACGCAAAAGAGAGAAATTCGAGATGTTCACACAGAGAA	2580
QY	2581	CGTGCCAAGTGTGTGGCCAGACAGCAGACAACATTAATCCCTCTAGCAATAGACGATTC	2640
Db	2581	CGTGCCAAGTGTGTGGCCAGACAGCAGACAACATTAATCCCTCTAGCAATAGACGATTC	2640
QY	2641	CGAAAGAGGTATATGTACTTAACATTAATCCCTCTGCGCTTTTGTTTTCAAACTTA	2700
Db	2641	CGAAAGAGGTATATGTACTTAACATTAATCCCTCTGCGCTTTTGTTTTCAAACTTA	2700
QY	2701	AGAGTAATCTGAATTAATCCGGTTTTGATTTCTTTGSCAAGCTGAGGAAGTGTCAAGCTTC	2760
Db	2701	AGAGTAATCTGAATTAATCCGGTTTTGATTTCTTTGSCAAGCTGAGGAAGTGTCAAGCTTC	2760
QY	2761	ATCGAGTTTCAAGAGAAAGAGATGAGGAGTTTGTGAAGAGAGGAGATCTGTATAAA	2820
Db	2761	ATCGAGTTTCAAGAGAAAGAGATGAGGAGTTTGTGAAGAGAGGAGATCTGTATAAA	2820
QY	2821	GATCAAGAGAAAGAGATGAAAGACATGAAGAAGGACATCAAGAGAGATATTTGATCTG	2880
Db	2821	GATCAAGAGAAAGAGATGAAAGACATGAAGAAGGACATCAAGAGAGATATTTGATCTG	2880
QY	2881	GAGAAAGAAATTTGATGAGGCTTTTGGAAACAGCTCATGTACAAGCATATGCGCTTCAACA	2940
Db	2881	GAGAAAGAAATTTGATGAGGCTTTTGGAAACAGCTCATGTACAAGCATATGCGCTTCAACA	2940
QY	2941	GATGATTTAGACAAAAGTCTGTATACACAAGACAAAGATCTAAGTTTCTTTGTCTTTTG	3000
Db	2941	GATGATTTAGACAAAAGTCTGTATACACAAGACAAAGATCTAAGTTTCTTTGTCTTTTG	3000
QY	3001	GTATGTCCGAAAGTAGAGATCTGAGACATTCATTTAATCTAGCAACAATCTTAAGA	3060

Db	3001	GTATGTCGGAAAGTAGAGAGATCTAGAGACTCCATTTAAATACAGCAAAATCTAAGCA	3060
QY	3061	GATTATAGATTATTATCTCTCCATTTTTAGTACCGGATCTAAGGAAGCATTTAAGTCTT	3120
Db	3061	GATTATGATTATTATTCTCCAAATTTTATAGACGGATCTAAGGAAGCATTTAAGTCTT	3120
QY	3121	GTGACATAAAACCAAGTTCCTTAGTATTTTGTTTTTTGTAAAAATTCATATGNAAG	3180
Db	3121	GTGACATAAAACCAAGTTCCTTAGTATTTTGTTTTTTGTAAAAATTCATATGNAAG	3180
QY	3181	TTAACAATATTACCAAAACGTACAGGTGAATCAACAGATGGCAATCAAAATCATGTTTTT	3240
Db	3181	TTAACAATATTACCAAAACGTACAGGTGAATCAACAGATGGCAATCAAAATCATGTTTTT	3240
QY	3241	AGATTTTTATATCTACAAAATATTGCGTCAAAAT	3275
Db	3241	AGATTTTTATATCTACAAAATATTGCGTCAAAAT	3275

LOCUS	AB025633	81365 bp	DNA	linear	PLN 27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:W0M1.				
ACCESSION	AB025633 BA000015				
VERSION	AB025633.2 GI:10178221				
KEYWORDS	.				
SOURCE	Arabidopsis thaliana (thale cress)				
ORGANISM	Arabidopsis thaliana				
REFERENCE	1 (sites)				
AUTHORS	Sato,S., Nakamura,Y., Kaneko,T., Katoh,T., Asamizu,E., Kotani,H. and Tabata,S.				
TITLE	Structural analysis of Arabidopsis thaliana chromosome 5. X.				
JOURNAL	Sequence features of the regions of 3,076,755 bp covered by sixty P1 and TAC clones				
MEDLINE	DNA Res. 7 (1), 31-63 (2000)				
PUBMED	20181125				
REFERENCE	10718197				
AUTHORS	2 (bases 1 to 81365)				
TITLE	Nakamura,Y.				
JOURNAL	Direct Submission				
COMMENT	Submitted (02-APR-1999) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:yasukazu@kazusa.or.jp, Tel:81-438-52-3935, Fax:81-438-52-3934)				
	On Sep 15, 2000 this sequence version replaced gi:4589439.				
	Address for correspondence: kase@kazusa.or.jp				
	For the latest information on annotation of this clone, please see http://www.kazusa.or.jp/kaos/cgi-bin/sgd_graph.cgi?c=W0M1				
	The software programs used to predict genes include: Graal (Informatics Group, Oak Ridge National Laboratory, http://compbio.ornl.gov/Graal-1.3/),				
	GENSCAN (Chris Burge, MIT, http://CCR-081.mit.edu/GENSCAN.html),				
	NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, http://www.cbs.dtu.dk/services/NetGene2/) and				
	SplicePredictor (Volker Brendel, Stanford University, http://gremli.ni.zool.jaist.ac.edu/cgi-bin/sp.cgi).				
	Genes encoding tRNAs are predicted by tRNAscan-SF (Sean Eddy, Washington University School of Medicine, St. Louis, http://genome.wustl.edu/eddy/tRNAscan-SF/).				
	This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.				
	The 5' clone is K19M13 and the 3' clone is W0M1.				
FEATURES	Location/Qualifiers				
SOURCE	1. 81365				
	/organism="Arabidopsis thaliana"				
	/mol_type="genomic DNA"				

/db xref="Columbia"
/db xref="taxon:3702"
/chromosome="5"
/clone="MQM1"
/clone_1db="Mitsui P1"
join(14922..3403,3474..3595,3681..3723,3777..3898,4042..4198,4445..4515,4601..4714,4888..4988,5300..5420)
/note="gb|AA01580.1
gene_id:MQM1.24
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97251.1"
/db xref="GI:8809710"
/translation="MHVPGISVDESFAHKKVDTPPKVQVLEPISTVVDAAQNS
YGAFFDPPSSNSPLSPVEEPSSFSBGDDPLGIEDLOISGEPYGHLOAG
YSINGTSCNFEWCHLEDGSVNYIDSSREKTSNAGAKPNLYLVADLYALIEV
OPDDRKRKGFAGKYRDLVLDVHTLFSADCFSEFFEMFALFICYLISGG
ACBDPMQSNIEKTIHTGHSYKSLAVGVPIWEATLSIKRGVSIKGISDITLAE
KFSASTVTTPROGPAELVYIGSDGSHSLADNGSPDLIGRDEIVTLRLFKRM
FFKALNFYHFRGCHKSGNL"
join(6317..6428,6508..6629,6714..6756,6841..6932,7121..7169,7396..7466,7550..7663,7785..7882,8112..8181)
/note="gb|AA01580.1
gene_id:MQM1.23
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97250.1"
/db xref="GI:8809709"
/translation="MCPGTSVYESPALNDADEFPPISTVVDNPKNIFYPDDGDAFL
PALENOISGEPYGHLOAGYSINGTSCNFEWCHLEDGSVNYIDGAKKNLYLT
ADVUGCLAEVOPLDNRKGLVYKPAENDKLACHPEMOSINDKTLHTGHSYV
SLAIGPHIMEATLSIERGYTKKNDNLITTEKBSATAVKIPEKPAELVYIGSD
GSEHCLRVNDEWPIRLCKEKVRLVKKVVEQKL"
complement(join(8642..8705,8794..8863,8978..9387,9584..10347))
/note="gene_id:MQM1.22
unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97249.1"
/db xref="GI:8809708"
/translation="MSMMKGSQSGWTAFDLKOROKGLESEVSGDPPPVSTVNAF
GVRGLRNRHPSKFSVILPSPRPALTENKDCNORGGCKRKPPLSLPNVS
HDLAFTLKEMNSWADNLRDVLSTEDPEMAFLKGVSSGKDEPSTISIBY
SSDRSEYRTEFTVTSVMAARSTFEDAGKDLNDSGSSFLVNASONEKFPDPI
SELDSIIQRLQSIPIEBEEDDLYLSHRDALKVMSASNSHSAANAFORVHSA
KOHSDKAREDMIAEKLNAAAKKIIGITNKDNDIKWLDLHGLHATAVVALORLOM
IEGHFTVRSVSPNRKSKNAALASASOEPPGRDLDEGMECORTSRLANSQVITG
IGKRGASLPLAVKTFPPDNRYRFPETRGVITVAPKFRHS"
complement(join(12417..13424)
/note="contains similarity to unknown protein
gb|AA07018.1
gene_id:MQM1.21"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97248.1"
/db xref="GI:8809707"
/translation="MATDSQPNOKLTPLKTRIALTVISMTDNAORPDGTINRFLR
LFPBRAPNPKPVNIYSTDSVDVQSDRLMFLRLTPHVSQDKIPVVFFHGGGAPFS
NADYDNVNGCFARFKLPAYIVSVYRLAPHRPAOYDGFALKEENHGSILPA
PMASRCPFAGDSAGNIANVARIRIGRPPSFTAKLIGTISIOBPGRERTRE
KOLYGAFLVSPDRDKMCKMGLNRDHAAYVGGPNVVISGDIPEPTVYVAGFDL
KDWRSYEMWLCKGKATLLEYNMFHAFIFPELPEAGQLIRIKDFDERVASIS
A"
join(14922..14981,15072..15167,15563..15666,15755..15891,16096..16178)
/note="gene_id:MQM1.20"
/codon_start=1
/evidence=not experimental
/product="50S ribosomal protein L24"
/protein_id="BAA97247.1"

/db xref="GI:8809706"
/translation="MGMKAAEKILRMKTLRGDNWMTIRGDKGNGTILKRVRSQNR
VTEGKMLIKKHIGKGFHEGGIFTEAPLHASNVOYDVPVTRGPKGVKYLIEDTK
VRVARGTSGSIIPEPELKIAPRPFPTAGKDPMEFWEQYDATTGKMPDL"
join(16823..16952,17143..17320,17412..17493,17675..17851,17943..18110,18199..18390)
/note="gene_id:MQM1.19"
/codon_start=1
/evidence=not experimental
/product="26S proteasome, non-ATPase regulatory subunit"
/protein_id="BAA97246.1"
/db xref="GI:8809705"
/translation="MERLORIFGAGGLGHSAPSPDLTSEQVYSSIALTKMLKHG
RAGVPEWNGMGEFVDEYTVRVYVDFPMOSGVSVAEDVHQTMDLXOTG
REPMVGVYHSHDFGCGTSLGVDINTQGFELANQAAVAVYDPIOSYKGVYDAFR
SINPQIMVGEERQTSNLGHNKRSIQALHGLNRHYSLAINRKNLEBKMLN
LHKKMTDGLTLRRPPTHKTEQVQEMSLAAKXKXVOEEDSLSPKALVNNGR
ODAKKLHEVSNLMSNIVQTLGMLDTVVF"
complement(join(18965..19043,19240..19298,19419..19536,19666..19739,19908..19961,20387..20530))
/note="contains similarity to unknown protein
emb|CAA2897.1
gene_id:MQM1.18"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97245.1"
/db xref="GI:8809704"
/translation="KDKMQALERKMLVMEVEDEQQAADSSISFMEIDLNRNAL
TKORPYGAICLSAGLTCTLSLVFNPVKGITLITGNLALSTGLFLIGPOKRV
TMDLDPARYATLVLASIIIALFCALVYRNKLLTLALITFTGLIWSLYIPRAR
TWSKIFMPCFDEP"
join(22672..23634,23708..23999,24111..24355,24458..24624,24715..24925)
/note="emb|CAB62356.1
gene_id:MQM1.17
similar to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97244.1"
/db xref="GI:8809703"
/translation="MSSRAGPMSKERNVOGYRBEVEQLVQSLAGTRLASSODGGEW
EVSKNKKNKPGNTSGTWSQNSNPPRMGGQOQSGSNVSGRGNVSGRGNNGRG
IQANISGRGALSRYKDNFVAPVPVRSPLGGMWQAGSQAHTVAQEPDVEDD
VONASRENDSDALDDSDDLSDVDSVSOSSHRSKONKMKFEPGSLDLSLEO
INPQOWHCPACONGPACIDWNLPLLAHATKARVRLHRELAFLYLOLMRG
ASVIFGERTYGGQGBDEKYEITVPMVITMTRLDKDNDKMGNOGLLYP
DKTEARASHSYGPGHRSVIMPESSATGVEARLRELEWGLDRIAMQKSM
FSGGVOLYGLATKQDLDI FNHSGKTRLRKELYSYOMVVKELQISSEDOQRLY
FKGSLKONKHAVLESLEINSEKLRATAEDNRIVROTKMOHEONREMDADRFF
MDSIKOIHERRDKENPFMLQOQERAKVVGQOQONINPSNDCKRAEVSSFLF
OEKEMEPFBERMLIKQEKKEKMDKKHHEIFLEKEFDALBQLMYKGLAHND
D"
join(25614..25838,26061..26207,26324..26438,26516..26610,26710..26813,27522..27765,27943..28036,28115..28305,28385..28604,28688..28855,28944..29092)
/note="gb|AAC97420.1
gene_id:MQM1.16
strong similarity to unknown protein"
/codon_start=1
/evidence=not experimental
/protein_id="BAA97243.1"
/db xref="GI:8809702"
/translation="MAPPGQTAAVAAGADGAPPOQGRGSGTISGIVRAVWFY
ASKFRPKQKMPDPTAPSQLMNTLPHKGSIDMWFLSLQEKNDKNDGPRALVMTEN
IPVAVTPESIRKSLITYPSKDLQNGSLVATIFPARSGFIDPTDPEIOPANSFR
THAVATYFPQKKKRRKSLSGSKDSESEPEVEKDGKSDKEEVPVEMISLMPN
VITNLVDYDFADLVEPTTGAVYPTIIFENEMWLDKRIPIVNEVSELPMLKSP
ISMMKQOLFQOVQSFOMORSYSGMIDGSDDELKVFLEGNPYLALITMFSMLHSF
DLFAPKNDIOFMNKNKSMGLSAKSVLNLPICOFVLEVLNDTSMITLASSGVGC
IEFWKIGKMRITVDSGMIPLRPHDRSSYSNKKVEYDILIKFLSVYLLLVIGT
SIYSLAVERHKSMTSLSSLSVTSVYMFIMCPDLFINYKLSVAHLPMRQMTYKF
LNTIIDLFAFVIMPIRLHRLSVFRDVIPLVLYORWYVVDKTRVNEFGGDEBT
AEKLLITERBEDKKTN"

CDS complement(join(29439..29576,29661..29885,29985..30268,
30341..30493,30579..30722,30956..31484))
/note="gene_id:MGM1.15

Query Match 100.0%; Score 3275; DB 8; Length 81365;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3275; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GACAAACAAACAAATTTAGCAGTATGTTGTAAGTAATTAATAGTGGAAACA 60
DB 21977 GACAAACAAACAAATTTAGCAGTATGTTGTAAGTAATTAATAGTGGAAACA 22036

QY 61 TTAAGTTAAGGAAAAAGAAAAAGTACAAAAATGAAAAATCAAACTGAAT 120
DB 22037 TTAAGTTAAGGAAAAAGAAAAAGTACAAAAATGAAAAATCAAACTGAAT 22096

QY 121 GAAAAATTTGAGTCCAGAAATCGAAAAACGAGCCGTTTAAAGCTTAATAGCTTC 180
DB 22097 GAAAAATTTGAGTCCAGAAATCGAAAAACGAGCCGTTTAAAGCTTAATAGCTTC 22156

QY 181 ATTGTCCTTCTTCGTCAGATTAATTTCTCCCGGAGTCCGACTCACTCTCAC 240
DB 22157 ATTGTCCTTCTTCGTCAGATTAATTTCTCCCGGAGTCCGACTCACTCTCAC 22216

QY 241 TCTCCGCGCTTTAACTTAAGTCTCCGTCCTTACTCTGTAGTTTTCGCTTAGAG 300
DB 22217 TCTCCGCGCTTTAACTTAAGTCTCCGTCCTTACTCTGTAGTTTTCGCTTAGAG 22276

QY 301 CTTCCGATCGCTCAACGCGATGATCTGTGTCGATTTCTCTTTTCTTCGCTGAAAA 360
DB 22277 CTTCCGATCGCTCAACGCGATGATCTGTGTCGATTTCTCTTTTCTTCGCTGAAAA 22336

QY 361 ATTGCCCTAATGTTCTGATTTTGAAGTTTGTGCTAGAGGTTACTTTTCCCATTA 420
DB 22337 ATTGCCCTAATGTTCTGATTTTGAAGTTTGTGCTAGAGGTTACTTTTCCCATTA 22396

QY 421 TTTTATAGTCTTAAGTAAACGATACCTGCGCTTACTGTTTGTTCATTTTGTGCT 480
DB 22397 TTTTATAGTCTTAAGTAAACGATACCTGCGCTTACTGTTTGTTCATTTTGTGCT 22456

QY 481 TTCACCGTTTATGTCGTGATCGAGATATTTGACTGTGAAAAATCCTTCGTTTGTG 540
DB 22457 TTCACCGTTTATGTCGTGATCGAGATATTTGACTGTGAAAAATCCTTCGTTTGTG 22516

QY 541 TTTGTTTATATTAATCGGATTTGATCTACCTTTTGTGCTTGAATTTTGTGACCT 600
DB 22517 TTTGTTTATATTAATCGGATTTGATCTACCTTTTGTGCTTGAATTTTGTGACCT 22576

QY 601 ATGCGTTTGTGCTGTTTATTAACCTTCACTGATGATTTTGAATTTTGTGAGCT 660
DB 22577 ATGCGTTTGTGCTGTTTATTAACCTTCACTGATGATTTTGAATTTTGTGAGCT 22636

QY 661 ACTGTGGGTTTCTTTGTTGCTATAGTGTGTAATAATGAGTTCTAGGGCTGTGCTCA 720
DB 22637 ACTGTGGGTTTCTTTGTTGCTATAGTGTGTAATAATGAGTTCTAGGGCTGTGCTCA 22696

QY 721 CTAAAGAAAAAGAGCTTCAAGGTGTGTAATAGGCTTGAAGTGAACAGTTGTTCAAG 780
DB 22697 CTAAAGAAAAAGAGCTTCAAGGTGTGTAATAGGCTTGAAGTGAACAGTTGTTCAAG 22756

QY 781 TGGCAGGAGCAGAGCTGGCTTCTTCACAAGATGATGAGGAGAGTGGAGGTCATTCCA 840
DB 22757 TGGCAGGAGCAGAGCTGGCTTCTTCACAAGATGATGAGGAGAGTGGAGGTCATTCCA 22816

QY 841 AGAAGAACAAAGAACAAACAGAAAACTTCTGAAAAAATCTTGAGTTTCTCAGAAATCGA 900
DB 22817 AGAAGAACAAAGAACAAACAGAAAACTTCTGAAAAAATCTTGAGTTTCTCAGAAATCGA 22876

QY 901 ATCTCTCTAAGCTTGGGGTGTGTGACAGACAGGAGAGAGTGAAGATCTTCTGGAGAG 960
DB 22877 ATCTCTCTAAGCTTGGGGTGTGTGACAGACAGGAGAGAGTGAAGATCTTCTGGAGAG 22936

QY 961 GAAACATGTATCCGGAGAGGTAAACGCAATGTTGGGGCATTCAGACTAATATCTGT 1020

DB 22937 GAAACATGTATCCGGAGAGGTAAACGCAATGTTGGGGCATTCAGACTAATATCTGT 22996

QY 1021 GTGGGGAGCAGAGCTTGAAGAGAAAGTAAATCAACTTTGGGCAACCCCACTGTAT 1080
DB 22997 GTGGGGAGCAGAGCTTGAAGAGAAAGTAAATCAACTTTGGGCAACCCCACTGTAT 23056

QY 1081 CTCGCCCTCTTGGAAAGAGATGGAATTGGCAGGCAAGAGAGTTCGTCTAGACA 1140
DB 23057 CTCGCCCTCTTGGAAAGAGATGGAATTGGCAGGCAAGAGAGTTCGTCTAGACA 23116

QY 1141 CAGCTGTGAGAGATTCTGACGTGAGAGATGATGATTAATGCTTCTGAGAAAGAGA 1200
DB 23117 CAGCTGTGAGAGATTCTGACGTGAGAGATGATGATTAATGCTTCTGAGAAAGAGA 23176

QY 1201 ATGATTTCCAGTCTTTGGATGATTTCTGATACGACCTTGAAGTGAATTAATGACTCGG 1260
DB 23177 ATGATTTCCAGTCTTTGGATGATTTCTGATACGACCTTGAAGTGAATTAATGACTCGG 23236

QY 1261 ATGTAGTCAAAAAGCCATGATGATCAAGAAAGCAATAGTGTCAAAAAGTCTTTG 1320
DB 23237 ATGTAGTCAAAAAGCCATGATGATCAAGAAAGCAATAGTGTCAAAAAGTCTTTG 23296

QY 1321 GCACTTGTAGTACCTTGTGATCGAGCAGTAAATGAACCAAGAGCAGTGGCATTTGC 1380
DB 23297 GCACTTGTAGTACCTTGTGATCGAGCAGTAAATGAACCAAGAGCAGTGGCATTTGC 23356

QY 1381 CAGCTTGTCAAGACGACCTGTGTCATGATGTTGTTAACTTGCACCTCTACTACTC 1440
DB 23357 CAGCTTGTCAAGACGACCTGTGTCATGATGTTGTTAACTTGCACCTCTACTACTC 23416

QY 1441 ATGGAGAGCAAAAAGAGCTAGGCGAGTTAAGCTCCATAGAAATGGCTGAAGTTTAA 1500
DB 23417 ATGGAGAGCAAAAAGAGCTAGGCGAGTTAAGCTCCATAGAAATGGCTGAAGTTTAA 23476

QY 1501 AAAAGATCTTAAGATGAGAGGCGCATCTGTCAATCTTGTGAGTGAATTTATGGCAGT 1560
DB 23477 AAAAGATCTTAAGATGAGAGGCGCATCTGTCAATCTTGTGAGTGAATTTATGGCAGT 23536

QY 1561 GGAAGGTTTGGGTGAGAGATGAAAAGATTAATGAATTTGCTGCTCCAAATGCTATCA 1620
DB 23537 GGAAGGTTTGGGTGAGAGATGAAAAGATTAATGAATTTGCTGCTCCAAATGCTATCA 23596

QY 1621 TCAATGAATCTAAGCTGATGATGAGATGATGATGATGATGATGATGATGATGATGAT 1680
DB 23597 TCAATGAATCTAAGCTGATGATGAGATGATGATGATGATGATGATGATGATGATGAT 23656

QY 1681 TTTCTTAATTTTCTCTTGCAATCTAATCTTGAATGTTCAATTTGATGCTCGGC 1740
DB 23657 TTTCTTAATTTTCTCTTGCAATCTAATCTTGAATGTTCAATTTGATGCTCGGC 23716

QY 1741 ATGGGCAACCAAGAGCTGTGGAATTAATCTGCAAGATGAGCTCTTAGAGCAGCCAT 1800
DB 23717 ATGGGCAACCAAGAGCTGTGGAATTAATCTGCAAGATGAGCTCTTAGAGCAGCCAT 23776

QY 1801 TCCATAGTCTCAAGGCGCATGCTGGATGAGTGTCTGATGATGATGATGATGATGATGAT 1860
DB 23777 TCCATAGTCTCAAGGCGCATGCTGGATGAGTGTCTGATGATGATGATGATGATGATGAT 23836

QY 1861 GGCATATTTGGAGGCGCAACGCTCCACCGGAGTGTAGCTGATGAGGTTGATAGATTT 1920
DB 23837 GGCATATTTGGAGGCGCAACGCTCCACCGGAGTGTAGCTGATGAGGTTGATAGATTT 23896

QY 1921 GCTTGGGCTCAAGAGCGATATGTTTCTGAGGTTGTCGCAACTGTATGCTTCTT 1980
DB 23897 GCTTGGGCTCAAGAGCGATATGTTTCTGAGGTTGTCGCAACTGTATGCTTCTT 23956

QY 1981 GCAACGAGCAAGATCTGACATATTCATCAACATCACTCAAGATTTCTCCCAAGA 2040
DB 23957 GCAACGAGCAAGATCTGACATATTCATCAACATCACTCAAGATTTCTCCCAAGA 24016

QY 2041 AATTTGATATATCTTTAGTTTGTTCATTTGATGAAATTTAAAGTTTGTGCTCGTGTAA 2100

Db	24017	AAATTGATATATGCTTTTACGTTTGTGTCATTGGAAATTTAAAGTTTGTGTGGCTGGTTAA	24076
Qy	2101	TGCATCTGTATGTATATATCTATGATTCATTAGGCAAAAACAAGCTGAAATTCGAGTTG	2160
Db	24077	TGCATCTGTATATGTATATCTATGATTCATTAGGCAAAAACAAGCTGAAATTCGAGTTG	24136
Qy	2161	AAATCTTACCAAGATGCGTGTGTTAAAGACCTGAGGAGATCTCTGAGGCAATCAGCAG	2250
Db	24137	AAATCTTACCAAGATGCGTGTGTTAAAGACCTGAGGAGATCTCTGAGGCAATCAGCAG	24196
Qy	2221	CTGAACCTACTTTAAGAAACAAGCTCTCAAAACAGAACAGCCAGCGCTCTGAGGAA	2280
Db	24197	CTGAACCTACTTTAAGAAACAAGCTCTCAAAACAGAACAGCCAGCGCTCTGAGGAA	24256
Qy	2281	TCTCTGAAAATTATGACGGAAGCTGCGTAGAACTGCAGAGATTAATCCGATCTGAGA	2340
Db	24257	TCTCTGAAAATTATGACGGAAGCTGCGTAGAACTGCAGAGATTAATCCGATCTGAGA	24316
Qy	2341	CAGAGAACTTAAGTCAGACATGAACAGAACAGGAAAGGTATGATTTTCTTGAAAAT	2400
Db	24317	CAGAGAACTTAAGTCAGACATGAACAGAACAGGAAAGGTATGATTTTCTTGAAAAT	24376
Qy	2401	CAACAACCTGACATTTTGTATTCCTATCTGATTCACATTTTGTATTAATATGCAACA	2460
Db	24377	CAACAACCTGACATTTTGTATTCCTATCTGATTCACATTTTGTATTAATATGCAACA	24436
Qy	2461	AAAACTGTGTGGTGTGAAGATGATGACAACAAGGTTTTTCATGATTCATCAAA	2520
Db	24437	AAAACTGTGTGGTGTGAAGATGATGACAACAAGGTTTTTCATGATTCATCAAA	24496
Qy	2521	CAGATCCATGAAGAAAGACACGCAAGAGAGAAATTTCCAGATGTTCCAGACACAGAA	2580
Db	24497	CAGATCCATGAAGAAAGACACGCAAGAGAGAAATTTCCAGATGTTCCAGACACAGAA	24556
Qy	2581	CGTGCCAAGGTTGTGGCCAGCAGCAGACAGAACATTAATCCCTCTAGCAATGACGATTGC	2640
Db	24557	CGTGCCAAGGTTGTGGCCAGCAGCAGCAGACAGAACATTAATCCCTCTAGCAATGACGATTGC	24616
Qy	2641	CGAAGAAGTATATGTACTAATCAATTAATCCCTGCGGTTTTGTTTTCAAACTTA	2700
Db	24617	CGAAGAAGTATATGTACTAATCAATTAATCCCTGCGGTTTTGTTTTCAAACTTA	24676
Qy	2701	AGAGTAACTGAATTAATCCGGTTTTGATTTCTTTGCAGAGCTGAGAAAGTCAAGCTTC	2760
Db	24677	AGAGTAACTGAATTAATCCGGTTTTGATTTCTTTGCAGAGCTGAGAAAGTCAAGCTTC	24736
Qy	2761	ATGAGGTTTCAAGAGAAAGATGAGAGAGTTTGTGAGAGAGAGGAGATCTGATTAAG	2820
Db	24737	ATGAGGTTTCAAGAGAAAGATGAGAGAGTTTGTGAGAGAGAGGAGATCTGATTAAG	24796
Qy	2821	GATCAAGAGAAAGATGAGAGACATGAAGAAGAGCATCAAGAGATATTGATCTG	2880
Db	24797	GATCAAGAGAAAGATGAGAGACATGAAGAAGAGCATCAAGAGATATTGATCTG	24856
Qy	2881	GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTACAAGCATGCGCTTCAACATGA	2940
Db	24857	GAGAAAGATTTGATGAGGCTTTGGAACAGCTCATGTACAAGCATGCGCTTCAACATGA	24916
Qy	2941	GATGATTGACAAAGCTGCTGTACACAAGACAAGACTAAGGATTTCTGTTGTTGCTTTTG	3000
Db	24917	GATGATTGACAAAGCTGCTGTGTACACAAGACAAGACTAAGGATTTCTGTTGTTGCTTTTG	24976
Qy	3001	GTAATGTCGAAAGTAGAGAGATCTGAGAGATCCATTTTAATTACTAGCAATCTTAAGA	3060
Db	24977	GTAATGTCGAAAGTAGAGAGATCTGAGAGATCCATTTTAATTACTAGCAATCTTAAGA	25036
Qy	3061	GATTATAGATTAATTAATCTCCAAATTTTATGAGCGATCTAAGAGACATTAAGTCTT	3120
Db	25037	GATTATAGATTAATTAATCTCCAAATTTTATGAGCGATCTAAGAGACATTAAGTCTT	25096
Qy	3121	GTAGACTAAAACAAGTTTCCCTTAGATTTTGTTTTTTGGTAAATTTCAATGAAG	3180
Db	25097	GTAGACTAAAACAAGTTTCCCTTAGATTTTGTTTTTTGGTAAATTTCAATGAAG	25156

[illegible]

Db 61 GAGGTGAACAGTTGGTTCAAGGTTGGCAAGGACGAGCTGGCTTCTTCAACAAGATGAT 120
Qy 816 GAGAGAGAGTGGAGGTCATTTTCCAGAGAAACAAGAACCAAGAGAACCTTCTGGA 875
Db 121 GAGAGAGAGTGGAGGTCATTTTCCAGAGAAACAAGAACCAAGAGAACCTTCTGGA 180
Qy 876 AAAAATCTGGGTTTCTGAGAAATTCGAATCTCTAGAGCTTGGGGTGGTCAAGCAAGG 935
Db 181 AAAATCTGGGTTTCTGAGAAATTCGAATCTCTAGAGCTTGGGGTGGTCAAGCAAGG 240
Qy 936 AGAGTAGCAACGATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGGCAATGT 995
Db 241 AGAGTAGCAACGATCTGGAGAGAGAAACAATGTATCCGGAGAGGTAAACGGCAATGT 300
Qy 996 CGGGGCTTCAAGCTAACATATCTGGTGGGAGACGAGCTTGGAGCAAAAATGATATAC 1055
Db 301 CGGGGCTTCAAGCTAACATATCTGGTGGGAGACGAGCTTGGAGCAAAAATGATATAC 360
Qy 1056 AACTTGTGGCACCCCACTGTATCTGGCTCTCTTGGAGAGGATGGAATTTGGCAG 1115
Db 361 AACTTGTGGCACCCCACTGTATCTGGCTCTCTTGGAGAGGATGGAATTTGGCAG 420
Qy 1116 GCAAGAGAGGTTCTGCTCAGCAACAAGCTGTGAGAGGTTTCTGACGTGAGAGATAT 1175
Db 421 GCAAGAGAGGTTCTGCTCAGCAACAAGCTGTGAGAGGTTTCTGACGTGAGAGATAT 480
Qy 1176 GTGATATATCTTCTGAGAGAGAGATGATTCGATCTTGGATGATTTGATGACAC 1235
Db 481 GTGATATATCTTCTGAGAGAGAGATGATTCGATCTTGGATGATTTGATGACAC 540
Qy 1236 CTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1295
Db 541 CTGCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Qy 1296 AATAGTGTCTTCAAAAAGTTCTTGGAGCTTGGATGATGATGATGATGATGATGAT 1355
Db 601 AATAGTGTCTTCAAAAAGTTCTTGGAGCTTGGATGATGATGATGATGATGATGAT 660
Qy 1356 GAAACACAGAGGAGTGGGATTTGTCAGCTTGTGACAAAGGAGCTGGTGCATGATGG 1415
Db 661 GAAACACAGAGGAGTGGGATTTGTCAGCTTGTGACAAAGGAGCTGGTGCATGATGG 720
Qy 1416 TATACCTGCAACCTCTACTAGCTCATGAGGACAAAGAGGATGAGGAGTAAAGCTC 1475
Db 721 TATACCTGCAACCTCTACTAGCTCATGAGGACAAAGAGGATGAGGAGTAAAGCTC 780
Qy 1476 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGATCTGCTAT 1535
Db 781 CATAGAGATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGATCTGCTAT 840
Qy 1536 CCTTGTGGTGAATTTATGAGGAGTGGAGAGGTTGGGTGAGATGAAGATTTATGA 1595
Db 841 CCTTGTGGTGAATTTATGAGGAGTGGAGAGGTTGGGTGAGATGAAGATTTATGA 900
Qy 1596 ATTGTCTGGCTCAATGATGATCATGATGATGATGATGATGATGATGATGATGATGAT 1655
Db 901 ATTGTCTGGCTCAATGATGATCATGATGATGATGATGATGATGATGATGATGATGAT 960
Qy 1656 AAGGTGAATTTCTTGTCTTTTACTTTTAAATTTTCTTGAATTTCTTACTGATCTTA 1715
Db 961 AAGGTGAATTTCTTGTCTTTTACTTTTAAATTTTCTTGAATTTCTTACTGATCTTA 1020
Qy 1716 GAATGTTTACTTGTAGTGGCTGGCATGGGCAACCAAGAGCTGTGATATCTTGCACA 1775
Db 1021 GAATGTTTACTTGTAGTGGCTGGCATGGGCAACCAAGAGCTGTGATATCTTGCACA 1080
Qy 1776 GTATGAGGCTCTTAAGAGAGCATTCCTATGATGATGATGATGATGATGATGATGAT 1835
Db 1081 GTATGAGGCTCTTAAGAGAGCATTCCTATGATGATGATGATGATGATGATGATGAT 1140
Qy 1836 TCTGATGTTTGAAGAGAGTGCACCTGGCTATTTGAGAGCGCAAGCTTCCACCGGAGTT 1895

Db 1141 TCTGATGTTTGAAGAGAGTGCACCTGGCTATTTGAGAGCCGCAAGCTTCCACCGGAGTT 1200
Qy 1896 AGCTGAGATGGGTTAGATAGATTTGCTTGGGTCAGAAACGAGATATGTTTTCTGAGG 1955
Db 1201 AGCTGAGATGGGTTAGATAGATTTGCTTGGGTCAGAAACGAGATATGTTTTCTGAGG 1260
Qy 1956 TGTTCGCAACTGATGAGCTTCTCTTGGCAAGAGCAAGATCTGGAACATATTCATCAACA 2015
Db 1261 TGTTCGCAACTGATGAGCTTCTCTTGGCAAGAGCAAGATCTGGAACATATTCATCAACA 1320
Qy 2016 CTCTCAAGTCTCTCCCAAGAAATTTGATATATGCTTTTATGTTTGTCAATGGAAT 2075
Db 1321 CTCTCAAGTCTCTCCCAAGAAATTTGATATATGCTTTTATGTTTGTCAATGGAAT 1380
Qy 2076 TTTAAATTTGTTGGTCCGTTTAAAGCATCTGTATATATATATATATATATATATAT 2135
Db 1381 TTTAAATTTGTTGGTCCGTTTAAAGCATCTGTATATATATATATATATATATATAT 1440
Qy 2136 CAAAACAAAGCTGAATTTGAGTTGAATCATACCAAGAGATGTTTAAAGAGCTGAG 2195
Db 1441 CAAAACAAAGCTGAATTTGAGTTGAATCATACCAAGAGATGTTTAAAGAGCTGAG 1500
Qy 2196 GAGATCTCTGAGAGCAATCAGAGCTGAACTTATTAAGAACAGCTCTCAAAACAGAA 2255
Db 1501 GAGATCTCTGAGAGCAATCAGAGCTGAACTTATTAAGAACAGCTCTCAAAACAGAA 1560
Qy 2256 CAAGCAGCCCAAGGCTTGAAGATCTCTGAGAAATTTATGAGAGAGAGCTGGGTAGAC 2315
Db 1561 CAAGCAGCCCAAGGCTTGAAGATCTCTGAGAAATTTATGAGAGAGAGCTGGGTAGAC 1620
Qy 2316 TGCAGAGATATATCGATCTGAGACAGAGACCTAAGATSCAGATATGACAGAACAGGA 2375
Db 1621 TGCAGAGATATATCGATCTGAGACAGAGACCTAAGATSCAGATATGACAGAACAGGA 1680
Qy 2376 AGAGTATGATTTTTCTTGAAGAAATCAAACTTGAATTTTATATCTTACTGATTTCA 2435
Db 1681 AGAGTATGATTTTTCTTGAAGAAATCAAACTTGAATTTTATATCTTACTGATTTCA 1740
Qy 2436 CATTTTGATATATTTGTCACAAAGAAACCTGTGGTGGTTGAAGTGAATGACACGGA 2495
Db 1741 CATTTTGATATATTTGTCACAAAGAAACCTGTGGTGGTTGAAGTGAATGACACGGA 1800
Qy 2496 CAGTTTTTCAATGATTTCAATCAACAGATCCATGAAAGAGAGACGCAAGAGAGGAA 2555
Db 1801 CAGTTTTTCAATGATTTCAATCAACAGATCCATGAAAGAGAGACGCAAGAGAGGAA 1860
Qy 2556 TTTGAGATGTTTGCACAGAGAGAGAGTGCACAGGTTGTTGGCCAGAGAGAGAGAT 2615
Db 1861 TTTGAGATGTTTGCACAGAGAGAGAGTGCACAGGTTGTTGGCCAGAGAGAGAGAT 1920
Qy 2616 TAATCCCTCTAGAGATGAGATTTGCCGAAAGAGGTATATGATTAATCAATATCCCT 2675
Db 1921 TAATCCCTCTAGAGATGAGATTTGCCGAAAGAGGTATATGATTAATCAATATCCCT 1980
Qy 2676 CTGGGCTTTTGTTTTCAAACTTAAGAGTGAATTAATTCGGTTTTGATTTCTTTCG 2735
Db 1981 CTGGGCTTTTGTTTTCAAACTTAAGAGTGAATTAATTCGGTTTTGATTTCTTTCG 2040
Qy 2736 CAGAGCTGAGAGAGTGCAGGCTTCAATGAGTTTCAAGAGAAAGAGATGAGAGAGTTTGT 2795
Db 2041 CAGAGCTGAGAGAGTGCAGGCTTCAATGAGTTTCAAGAGAAAGAGATGAGAGAGTTTGT 2100
Qy 2796 GGAAGAGAGGAGATGCTGATTAAGATCAAGAGAAAGATGGAAGACATGGAAGAGAG 2855
Db 2101 GGAAGAGAGGAGATGCTGATTAAGATCAAGAGAAAGATGGAAGACATGGAAGAGAG 2160
Qy 2856 GCATCAGAGAGATTTTGTATCTGAGAGAAAGATTTGATGAGGCTTTGGAACAGCTCAT 2915
Db 2161 GCATCAGAGAGATTTTGTATCTGAGAGAAAGATTTGATGAGGCTTTGGAACAGCTCAT 2220
Qy 2916 GTTCAAGATGGCTTCAACATGAGATGATGA 2949
Db 2221 GTTCAAGATGGCTTCAACATGAGATGATGA 2254

RESULT 4
BT002944
LOCUS
DEFINITION
Arabidopsis thaliana clone RAF14-93-K05 (R20243) linear PLN 15-JAN-2003
(At5g23570) mRNA, complete cds.
ACCESSION
BT002944
VERSION
BT002944.1 GI:27754622
KEYWORDS
FLI CDNA.
SOURCE
Arabidopsis thaliana (thale cress)
Arabidopsis thaliana
Bukerjocla; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 2162)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 2162)
Yamada, K., Chan, M.M., Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M.,
Onodera, C.S., Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C.,
Yu, G., Yuan, S., Carninci, P., Chen, H., Cheuk, R., Hayashizaki, Y.,
Ishida, J., Jones, T., Kamiya, A., Kawai, J., Kim, C.J., Narusaka, M.,
Nguyen, M., Palm, C.J., Sakurai, T., Satou, M., Seki, M., Shim, P.,
Southwick, A., Tripp, M.G., Wu, T., Shinozaki, K., Davis, R.W.,
Ecker, J.R. and Theologis, A.
Direct Submission
Submitted (15-JUN-2003) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA: 'RIKEN
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Ishida, J.,
Hayashizaki, Y. and Shinozaki, K.).

The Salk, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Yamada, K., Chan, M.M.,
Chang, C.H., Dale, J.M., Hsuan, V.W., Lee, J.M., Onodera, C.S.,
Quach, H.L., Tang, C.C., Toriumi, M., Wong, C., Wu, H.C., Yu, G.,
Yuan, S., Chen, H., Cheuk, R., Jones, T., Kim, C.J., Nguyen, M.,
Palm, C.J., Shim, P., Southwick, A., Tripp, M.G., Wu, T., Davis, R.W.,
Ecker, J.R. and Theologis, A.

Yamada, K. (SSP/PGEC) and Seki, M. (RIKEN GSC) contributed equally to
this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC)
contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome
submitted to Genbank.
FEATURES
source
1. 2162
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/chromosome="5"
/clone="RAFL14-93-K05 (R20243)"
/ecotype="Columbia"
/note="This clone is in a modified pBluescript vector2
(lambda ps) as a BamHI/XhoI insert."
1. 2162
/gene="At5g23570"
1. 91
/gene="At5g23570"
92. 1969
/gene="At5g23570"
/codon_start=1

ORIGIN
Query Match 29.7%; Score 972; DB 8; Length 2162;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 972; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
misc_difference 2166
/gene="At5g23570"
/note="not present in genomic sequence"
3' UTR
1970. 2162
/gene="At5g23570"
/evidence=experimental
/product="unknown protein"
/protein_id="AA022757.1"
/db_xref="GI:27754623"
/translation="MSSRAGPMKSKENNGGVRPEYEOI VOGLACRRLASQDDGSEW
EVISKNNKRRGTSKRTVNSQNSNPPRMGGQOQSRGNSVSGRGNVGKRGNGRG
IQANISGRGALSRKTDNNFVAPVPVSRPPLRGKMWQARGSAGHTVAQVQEPVDEVD
VDNASEENDSDALDDSDDLASDDSDVSQSHGSRKQNKWFKFQSLSLSTEQ
INBPORQPCACONGPGAIDWNLHLHARTKARRVRLKRLAEVLEKLOWRG
ASVIPCGEIYGQWKLGEDEKDYELVWPPMVIIMTRLDKNDNDKMLGNGQLELYF
DKYEALRAHRSYVPOGRGASVYMEFSATGVLAEARHRELAEMLGDLRIANGCRSM
FSGGVRLVYGFLLTRDILFNHSGKTRLEFKLSVQEMTKELROISNQOLNT
FRNKLKQNKHAKVLEESI FMSBKRLAEADRIVROTKRQHEQNRKEMADHDEF
WDSIKQIHRDRDKENFEMLDQOQERAKVVGQOQINPSSDDCKRAEVSSTLEF
QEKEMEEFVEEREMLIKDKQKEMDKKHBEIIPLEKEPDEALQIWMYKHGLNHED
D"
1970. 2162
/gene="At5g23570"
/note="not present in genomic sequence"
QY 687 GTTGTAAATGAGTTCTAGGGCTGCTCAATGCTTAGAAGAAAGAAACGTTCAAGGTGTGT 746
DB 83 GTTGTAAATGAGTTCTAGGGCTGCTCAATGCTTAGAAGAAAGAAACGTTCAAGGTGTGT 142
QY 747 TATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAAGATGCTTCTTCA 806
DB 143 TATAGGCTGAGGTTGAACAGTTGTTCAAGTTTGGCAGGAGCAAGATGCTTCTTCA 202
QY 807 CAAGATGATGAGGAGAGAGTGGAGGTCATTTCAAGAAAGAAACAACCAAGAAAC 866
DB 203 CAAGATGATGAGGAGAGAGTGGAGGTCATTTCAAGAAAGAAACAACCAAGAAAC 262
QY 867 ACTTCTGAAAAAATCTGGGTTTCTCAAGATTGGAATTCCTCTTAGAGCTTGGGGTGTGAG 926
DB 263 ACTTCTGAAAAAATCTGGGTTTCTCAAGATTGGAATTCCTCTTAGAGCTTGGGGTGTGAG 322
QY 927 CAGCAAGGAGAGGATGACAAAGTATCTGGAGAGAAACAATGTTCCGGGAGAGAGTAAAC 986
DB 323 CAGCAAGGAGAGGATGACAAAGTATCTGGAGAGAAACAATGTTCCGGGAGAGAGTAAAC 382
QY 987 GGCATGATGTCGGGGGATTCACATCAATCTGCTGGGGGAGCAGACCTTGAGCAGAAAG 1046
DB 383 GGCATGATGTCGGGGGATTCACATCAATCTGCTGGGGGAGCAGACCTTGAGCAGAAAG 442
QY 1047 TATGATTAACAATTTTGTGGCACCCCACTGATCTGCGCTCTTTGGAAGAGAGATGG 1106
DB 443 TATGATTAACAATTTTGTGGCACCCCACTGATCTGCGCTCTTTGGAAGAGAGATGG 502
QY 1107 AATTGGCAGGCAAGAGAGGTTCTGTCAGCACAAGCTGTGACAGAGTTTCTGACGTG 1166
DB 503 AATTGGCAGGCAAGAGAGGTTCTGTCAGCACAAGCTGTGACAGAGTTTCTGACGTG 562
QY 1167 GAGGATGATGTGATTAATGCTTCTGAGAGAGAAATGATTCGATCTTTGATGATATCT 1226
DB 563 GAGGATGATGTGATTAATGCTTCTGAGAGAGAAATGATTCGATCTTTGATGATATCT 622
QY 1227 GATGAGCAGCTTGGCAAGAGATGATTAATGATCTGATGATGATCAAAAGAGCATGATCA 1286
DB 623 GATGAGCAGCTTGGCAAGAGATGATTAATGATCTGATGATGATCAAAAGAGCATGATCA 682
QY 1287 CGAAGAGCAATTAATGATGATCAAAAGATCTTTGGCAGCTTGGATGATGATGATGATG 1346
DB 683 CGAAGAGCAATTAATGATGATCAAAAGATCTTTGGCAGCTTGGATGATGATGATGATG 742
QY 1347 CAGATTAATGAACACAGAGGAGTGGATTTGTCAGCTTGTCAAGACGATCTGTGTC 1406

Db 743 CAGATTAATGACCAAGAGGAGGAGTGGCATTTGTCCAGCTTGTCAAGAACGAGCTTGTC 802
Qy 1407 ATGATTTGGTATTAACCTGACCTCTCTACTAGCTCATGAGAGCAAAAAGAGCTAGGCGA 1466
Db 803 ATGATTTGGTATTAACCTGACCTCTCTACTAGCTCATGAGAGCAAAAAGAGCTAGGCGA 862
Qy 1467 GTTAAGCTCCATGAGAAATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCA 1526
Db 863 GTTAAGCTCCATGAGAAATTTGGCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCGCA 922
Qy 1527 TCTCTCATTTCTTGTGTGAGATTTATTTGAGAGAGGAGGTTTGGTGAAGATGAAGAAG 1586
Db 923 TCTCTCATTTCTTGTGTGAGATTTATTTGAGAGAGGAGGTTTGGTGAAGATGAAGAAG 982
Qy 1587 GATTATGAAATTTGCTGGCTCCATGAGTGCATCATCATGATTAAGTCTGATTAAGGAC 1646
Db 983 GATTATGAAATTTGCTGGCTCCATGAGTGCATCATCATGATTAAGTCTGATTAAGGAC 1042
Qy 1647 GATTAACGATTAAG 1658
Db 1043 GATTAACGATTAAG 1054

RESULT 5
AX078761 1878 bp DNA linear PAT 22-FEB-2001
-LOCUS AX078761
DEFINITION Sequence 2 from Patent WO0105951.
ACCESSION AX078761
VERSION AX078761.1 GI:13158380
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosid II; Brassicales; Brassicaceae; Arabidopsi.

REFERENCE
AUTHORS Beclin, C., Elmayan, T. and Vaucheret, H.
TITLE Novel sg83 plant gene and use thereof
JOURNAL Patent: WO 0105951-A 2 25-JAN-2001;
AVENUS CROSCIENCE S.A. (FR) ; INSTITUT NATIONAL DE LA RECHERCHE
AGRONOMIQUE (FR)

FEATURES
source
1. .1878
/organism="Arabidopsis thaliana"
/mol_type="unassigned DNA"
/db_xref="taxon:3702"
1. .1878
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC32419.1"
/db_xref="GI:13158381"
/translation="MSRAGPMSKERKVGQYRPEVQLVQGLAGTRLASSODDGGEM
EVLISKNNKRPNGTSGKTWVSQNSNPRPAGQGOQGSNVSGKNNVSGKNGNGG
IQANISGRGRLSRKYDNNFVAPVSRPLEGGMNQAQSGSKQTAQVDFPVEDD
VDNASEEDSDALDDDDLDSDVSDVSSQSHGRKNNKPKRFGSLDLSLSD
INBPOROMPCAPONGPAGIDWYHLHLHARTKGAIRKYLKLEVLKQLQMG
ASYIPCEITGQCKGLGEDEKDEYIIVPPIVITMTLDDDDNDKMLGKNOLELTF
DKYELAKHRSYGRQGHKMSVLAFESSATYLEAERLHELAMEGLDRLAWQKRM
FKGVRQLYGFATKQDLDFNHSQKRTLKEPLKSYQMVVYELAIQIBEDQOLNY
FKNKLSKONKHAFLERLELIMSKEKRLTIEDNRIYVQRTQMOHEORREMDADRFP
MDSIKOHERDAKEENFEMLOOQERAKVVGQOQONIPSSNDRCRAREVSPTEF
OEKEMEFVEREMILIDQEKKEKMDKKRHHBELFDEKEFDEALBQMLYKGLHND
D"

ORIGIN

Query Match 29.4%; Score 963; DB 6; Length 1878;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 963; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 696 ATGAGTTCTAGGGCTGTGCTCAATGTCTTAAGAAAGAGCTTCAGGGTGTATAGGCTT 755
Db 1 ATGAGTTCTAGGGCTGTGCTCAATGTCTTAAGAAAGAGCTTCAGGGTGTATAGGCTT 60

Qy 756 GAGGTGAAAGATTGTTCAAGTTTGGCAGAGGACGAGACTGGCTTCTTCAAGATGAT 815
Db 61 GAGGTGAAAGATTGTTCAAGTTTGGCAGAGGACGAGACTGGCTTCTTCAAGATGAT 120
Qy 816 GAGAGAGAGTGGAGGTCATTTCCAGAGAAACAAGAACCAAGAGAAACATCTTGA 875
Db 121 GAGAGAGAGTGGAGGTCATTTCCAGAGAAACAAGAACCAAGAGAAACATCTTGA 180
Qy 876 AAAAATTGGTTTCTCAAGATTCGAATCTTCTTGAAGTGGGTGTGACACAGAGG 935
Db 181 AAAAATTGGTTTCTCAAGATTCGAATCTTCTTGAAGTGGGTGTGACACAGAGG 240
Qy 936 AGAGTGAACACATCTGAGAGAGAAACAATGTATCCGAGAGAGGTAACCGCAATGAT 995
Db 241 AGAGTGAACACATCTGAGAGAGAAACAATGTATCCGAGAGAGGTAACCGCAATGAT 300
Qy 996 CGGGGCAATTCAGCTAACATATCTGATCGGGGAGAGAGCTTGAACAGAAAGTATGATAC 1055
Db 301 CGGGGCAATTCAGCTAACATATCTGATCGGGGAGAGAGCTTGAACAGAAAGTATGATAC 360
Qy 1056 AACTTTGTGGCACTCCCACTGTATCTGCTCTCTTGGAGAGAGATGATGACG 1115
Db 361 AACTTTGTGGCACTCCCACTGTATCTGCTCTCTTGGAGAGAGATGATGATGACG 420
Qy 1116 GCAAGAGAGAGTCTGCTCAGACACAGAGCTGAGAGAGTTTCTGACGTGAGAGATGAT 1175
Db 421 GCAAGAGAGAGTCTGCTCAGACACAGAGCTGAGAGAGTTTCTGACGTGAGAGATGAT 480
Qy 1176 GTGATATATGCTCTGAGAGAGAGATATTCGATGATGATGATGATGATGATGATGAT 1235
Db 481 GTGATATATGCTCTGAGAGAGAGATATTCGATGATGATGATGATGATGATGATGAT 540
Qy 1236 CTTGCAAGTATGATTTATGACTCGAGATGATGATCAAAAGAGCCATGATGATCAAGAGCAG 1295
Db 541 CTTGCAAGTATGATTTATGACTCGAGATGATGATCAAAAGAGCCATGATGATCAAGAGCAG 600
Qy 1296 AATAAGAGTTCAAAAAGTTCTTTGGCAGAGTGGATGATGATGATGATGATGATGATGAT 1355
Db 601 AATAAGAGTTCAAAAAGTTCTTTGGCAGAGTGGATGATGATGATGATGATGATGATGAT 660
Qy 1356 GAAACACAGAGAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1415
Db 661 GAAACACAGAGAGAGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Qy 1416 TATTAACCTGACCTCTACTAGCTCATGAGAGACAAAGAGAGCTAGGCAATTAAGCTC 1475
Db 721 TATTAACCTGACCTCTACTAGCTCATGAGAGACAAAGAGAGCTAGGCAATTAAGCTC 780
Qy 1476 CATTAAGAAATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGATCTGAT 1535
Db 781 CATTAAGAAATTTGGCTGAAGTTTGAAGAAAGATCTACAGATGAGAGGCGATCTGAT 840
Qy 1536 CTTTGTGTGAGATTTATGAGGAGTGAAGGTTTGGTGAAGATGAAGAAAGATTAAGAA 1595
Db 841 CTTTGTGTGAGATTTATGAGGAGTGAAGGTTTGGTGAAGATGAAGAAAGATTAAGAA 900
Qy 1596 ATTGTCTGGCTTCAATGATGATCATGATGATGATGATGATGATGATGATGATGATGAT 1655
Db 901 ATTGTCTGGCTTCAATGATGATCATGATGATGATGATGATGATGATGATGATGATGAT 960
Qy 1656 AAG 1658
Db 961 AAG 963

RESULT 6
BT004380 1909 bp mRNA linear PLN 14-FEB-2003
LOCUS BT004380
DEFINITION Arabidopsis thaliana clone U20243 unknown protein (At5g23570) mRNA,
complete cds.
ACCESSION BT004380
VERSION BT004380.1 GI:28393932

Db 901 ATTGTGGCTCCCAATGATCATCATGATACTAGCTGATTAAGACGATTAACGAT 960
Qy 1656 AAG 1658
Db 961 AAG 963

RESULT 7
ATHS28171
LOCUS 650 bp DNA linear PLN 29-MAR-2003
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone 157C08.
ACCESSION AJ528171.1 GI:26796431
VERSION 1
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., Dehose, R., Pelletier, G., Lepointec, L., Caboche, M. and Lecharny, A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
MEDLINE 12446565
PUBMED 1246565
REFERENCE 2 (bases 1 to 650)
AUTHORS Balzerque, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.inbio.gen.fr>).
FEATURES
source
1. .650
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/cultivar="KassilLewskiJa"
/db_xref="taxon:3702"
/clone="157C08"
/clone_1fb="Arabidopsis thaliana T-DNA insertion lines"
1. .650
/note="T-DNA flanking sequence
left border"

ORIGIN
Query Match 16.5%; Score 541; DB 8; Length 650;
Best Local Similarity 99.8%; Pred. No. 1.9e-296;
Matches 591; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1428 CCTCTACTACTGCTATGAGAGCAAAAGAGCTTGGGAGTTAAGCTCCATAGAGATTG 1487
Db 1 CCTCTACTACTGCTATGAGAGCAAAAGAGCTTGGGAGTTAAGCTCCATAGAGATTG 60
Qy 1488 GCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCCGATCTGTCATTCTCTTGAGGAG 1547
Db 61 GCTGAAGTTTAAAGAAAGATCTACAGATGAGAGCCGATCTGTCATTCTCTTGAGGAG 120
Qy 1548 ATTATGGAGCAGTGAAGGGGTTGGGTGAGAGTAAAGATTATGAATGTGTGACCT 1607

Db 121 ATTATGGAGCAGTGAAGGGGTTGGGTGAGAGTAAAGATTATGAATGTGTGACCT 180
Qy 1608 CCATGGCTCATTCAATCANGAATACAGCTGAGTAAGACGATTAACGATTAAGTGAATTTC 1667
Db 181 CCATGGCTCATTCAATCANGAATACAGCTGAGTAAGACGATTAAGTGAATTTC 240
Qy 1668 TTCTGTCTTTTACTCTCTTAAATTTTCTCTTGTGATCTAGATCTTGAATGTACATT 1727
Db 241 TTCTGTCTTTTACTCTCTTAAATTTTCTCTTGTGATCTAGATCTTGAATGTACATT 300
Qy 1728 GTATGCTCTGCGATGCGACCAACCAAGAGCTGCGAATCTTGACCAAGTATGAGGCTCT 1787
Db 301 GTATGCTCTGCGATGCGACCAACCAAGAGCTGCGAATCTTGACCAAGTATGAGGCTCT 360
Qy 1788 TAGAGCAGCATTTCTCATAGTCCACAGGCGCATTCGGAGATGATGATGATGATGATG 1847
Db 361 TAGAGCAGCATTTCTCATAGTCCACAGGCGCATTCGGAGATGATGATGATGATGATG 420
Qy 1848 GAGCAGTGCACACTGCTATTGAGAGCGCCAGACCTCCACCGGAGTTAGCTGAGATGAG 1907
Db 421 GAGCAGTGCACACTGCTATTGAGAGCGCCAGACCTCCACCGGAGTTAGCTGAGATGAG 480
Qy 1908 GTTAGATGAATGATGCTGCGGCTCAGAAAGCCAGATGTTTCTGAGAGTGTGCGCACT 1967
Db 481 GTTAGATGAATGATGCTGCGGCTCAGAAAGCCAGATGTTTCTGAGAGTGTGCGCACT 540
Qy 1968 GTATGCTCTGCGATGCGACCAACCAAGATCTGACATTTCAATCAACACTCT 2019
Db 541 GTATGCTCTGCGATGCGACCAACCAAGATCTGACATTTCAATCAACACTCT 592

RESULT 8
ATHS27741
LOCUS 45 bp DNA linear PLN 29-MAR-2003
DEFINITION Arabidopsis thaliana T-DNA flanking sequence, left border, clone 16F03.
ACCESSION AJ527741.1 GI:26796001
VERSION 1
KEYWORDS left border; T-DNA flanking sequence.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1
AUTHORS Brunaud, V., Balzerque, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., Dehose, R., Pelletier, G., Lepointec, L., Caboche, M. and Lecharny, A.
TITLE T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites
JOURNAL EMBO Rep. 3 (12), 1152-1157 (2002)
MEDLINE 12446565
PUBMED 1246565
REFERENCE 2 (bases 1 to 45)
AUTHORS Balzerque, S.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-2002) Balzerque S., UMRGV, INRA/CNRS, 2 rue Gaston Cremieux, 91057 Evry cedex, FRANCE
COMMENT PCR was performed on DNA from transformants of Arabidopsis thaliana plants from INRA (Versailles). The DNA fragment(s) resulting from the PCR were directly sequenced from the left or the right border to determine the genomic sequence flanking the insertion. T-DNA derived sequences were removed. Information to order the corresponding mutant line and a link to a database providing a graphical display of the insertion site are available at <http://dbsgap.versailles.inra.fr/publiclines/>. This sequence has been generated in the framework of the French plant genomics program 'Genoplante' (<http://www.genoplante.com> and <http://genoplante-info.inbio.gen.fr>).
FEATURES
source
1. .45
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"

```

misc_feature
/cultivar="Massillweskija"
/db_xref="taxon:3702"
/clone="146F03"
/clone.lib="Arabidopsis thaliana T-DNA insertion lines"
1.45
/notes="T-DNA flanking sequence
left border"

ORIGIN
Query Match 1.4%; Score 45; DB 8; Length 45;
Best Local Similarity 100.0%; Pred. No. 5.2e-13;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2033 CCCAAGAAATTGATATATGCTTATGTTTGTCTATTCGATTT 2077
1 CCCAAGAAATTGATATATGCTTATGTTTGTCTATTCGATTT 45

RESULT 9
AC102284/c
LOCUS
DEFINITION
Mus musculus clone RP24-329L16, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
AC102284 GI:24431710
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
MUS musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 170765)
Birren, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-329L16
Unpublished
2 (bases 1 to 170765)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,
Chopel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatord, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamacz, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marguis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McNeeters, R., Meltrin, J.,
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, U., Nguyen, C.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 170765)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Chopel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hateg, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R.,
Lindblad-Ton, K., Liu, G., Maclean, C., Macdonald, P., Major, J.,
Matthews, C., McCarthy, M., Meltrin, J., Meneus, L., Mihova, T.,
Mlenga, V., Murphy, T., Naylor, U., Nguyen, C., Nicoll, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,

```

TITLE
 JOURNAL
 COMMENT
 Submitted (31-OCT-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Oct 31, 2002 this sequence version replaced gi:22381142.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WITR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L18299
 Center clone name: 329 L 16
 ----- Summary Statistics
 Sequencing vector: Plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 169312 bases at least Q40
 Consensus quality: 170102 bases at least Q30
 Consensus quality: 170295 bases at least Q20
 Insert size: 16300; agarose-fp
 Insert size: 170365; sum-of-contigs
 Quality coverage: 17.6 in Q20 bases; agarose-fp
 Quality coverage: 16.8 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. The true order in this sequence record is
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

```

1 76853: contig of 76853 bp in length
* 76854 76953: gap of 100 bp
* 76954 78949: contig of 1996 bp in length
* 78950 79049: gap of 100 bp
* 79050 92772: contig of 13723 bp in length
* 92773 92872: gap of 100 bp
* 92873 150873: contig of 58001 bp in length
* 150874 150973: gap of 100 bp
* 150974 170765: contig of 19792 bp in length.
Location/Qualifiers
1. 170765
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-329L16"
/clone.lib="RPCT-24 Male Mouse BAC"
1. 76853
/notes="assembly_fragment
clone_end:SP6
vector_side:left"
76954. 78949
/notes="assembly_fragment"
79050. 92772
/notes="assembly_fragment"
92873. 150873
/notes="assembly_fragment"
150974. 170765
/notes="assembly_fragment
vector_side:right"

```

ORIGIN

DEFINITION Zebrafish DNA sequence from clone BUSM1-132M23 in linkage group 7, complete sequence.

ACCESSION AL845510

VERSION AL845510.8 GI:28172206

KEYWORDS HTG.

SOURCE Danio rerio (zebrafish)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 101241)

AUTHORS Bates, K.

JOURNAL Direct Submission
Submitted (29-JUN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
On Jan 30, 2003 this sequence version replaced gi:27801659.

COMMENT -----
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep

Repeat names beginning 'dr' were identified by the Repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'drr', were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see http://www/Projects/D_rerio/fishmask.shtml

BUSM1-132M23 is from a zebrafish PAC library

FEATURES VECTORS: pCYPAC-6.

SOURCE Location/Qualifiers

1..101241

 /organism="Danio rerio"

 /mol_type="genomic DNA"

 /db_xref="taxon:7955"

 /clone="BUSM1-132M23"

 /clone_lib="BUSM1"

ORIGIN

Query Match 0.7%; Score 23; DB 5; Length 101241;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 96 AATGTAACCAATCAACTGA 118
|||||
Db 39751 AATGTAACCAATCAACTGA 39729
|||||

RESULT 13

AC107028/c AC107028 126716 bp DNA linear PRI 29-MAR-2003

LOCUS Homo sapiens 3 BAC RP11-547K2 (Roswell Park Cancer Institute Human BAC library) complete sequence.

AC107028 AC107028 5 GI:28006935

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 126716)

AUTHORS Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Alshrocks, S.L., Amaral, H.C., Are, J.R., Ayala, M., Banke, T., Barbieri, J., Benton, J., Bimge, K., Blankenburg, K., Bonini, N.P., Bouck, J., Bowe, S., Brice, M., Brown, B., Brown, M., Bryant, N.P., Buha, J., Burck, P., Burkett, C., Butrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denu, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durkin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falle, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., He, X., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulys, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratoch, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichte, O., Lien, C., Liu, J., Liu, W., Louie, R., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhinney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabac, K., Moore, S., Morgan, M., Moorish, T., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogun, M., Okunolu, G., Ogunye, N., Oviedo, R., Pace, A., Payton, B., Peery, D., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojupokan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shooshitari, N., Sison, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svalik, A., Taber, P., Tameris, A., Tameris, K., Tang, H., Taney, J., Taylor, C., Taylor, T., Telford, B., Thomas, R., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Naylor, S.L., Weinstein, G. and Gibbs, R.

TITLE Direct Submission

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 126716)

AUTHORS Worley, K.C.

JOURNAL Direct Submission

Submitted (14-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 3 (bases 1 to 126716)

AUTHORS Worley, K.C.

JOURNAL Direct Submission

Submitted (29-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 4 (bases 1 to 126716)

AUTHORS Worley, K.C.

JOURNAL Direct Submission

Submitted (30-APR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 5 (bases 1 to 126716)

AUTHORS Worley, K.C.

JOURNAL Direct Submission

Submitted (02-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 6 (bases 1 to 126716)
 AUTHORS Morley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 7 (bases 1 to 126716)
 AUTHORS Morley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUN-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 8 (bases 1 to 126716)
 AUTHORS Morley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (15-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

REFERENCE 9 (bases 1 to 126716)
 AUTHORS Morley,K.C.
 TITLE Direct Submission
 JOURNAL Submitted (29-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jan 29, 2003 this sequence version replaced gi:20340431.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.htm>.

FEATURES
 source Location/Qualifiers
 1..126716
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="3"
 /clone="RP11-547K2"
 /complement(1..3011)
 /note="overlaps bases 1..3011 of clone AC026305"
 /function="clone overlap"
 repeat_region complement(28..209)

repeat_region /rpt_family="MER34"
 219..973
 /rpt_family="LIM4"
 repeat_region 1409..1455
 /rpt_family="AluY"
 repeat_region 1456..1754
 /rpt_family="AluY"
 repeat_region 1762..1998
 /rpt_family="AluSC"
 repeat_region 2145..2195
 /rpt_family="AT_rich"
 repeat_region 2612..2921
 /rpt_family="AluSX"
 repeat_region 3493..3520
 /rpt_family="AT_rich"
 repeat_region 3573..3598
 /rpt_family="AT_rich"
 repeat_region 5930..5955
 /rpt_family="AT_rich"
 repeat_region complement(6168..6478)
 /rpt_family="AluSP"
 repeat_region 6945..6984
 /rpt_family="(TTTA)n"
 repeat_region complement(6986..7276)
 /rpt_family="AluSX"
 repeat_region 11796..11975
 /rpt_family="MERSA"
 repeat_region complement(13903..13967)
 /rpt_family="L2"
 repeat_region 14192..14235
 /rpt_family="MIR"
 repeat_region 16052..16078
 /rpt_family="AT_rich"
 repeat_region complement(17082..17461)
 /rpt_family="MSTD"
 repeat_region complement(17583..17934)
 /rpt_family="MER21B"
 repeat_region 17944..17997
 /rpt_family="(TTATA)n"
 repeat_region complement(18003..18257)
 /rpt_family="AluB"
 repeat_region complement(18261..18557)
 /rpt_family="MER21B"
 repeat_region 19148..19215
 /rpt_family="AT_rich"
 repeat_region 19234..19342
 /rpt_family="FLAM_C"
 repeat_region 19377..19428
 /rpt_family="A-rich"
 /rpt_family="A-rich"

Query Match 0.7%; Score 23; DB 9; Length 126716;
 Best Local Similarity 100.0%; Pred. No. 1.9;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1673 TCTTTACTCTCTTAATTTTCT 1695
 Db 47492 TCTTTACTCTCTTAATTTTCT 47470

RESULT 14
 BX005128/c 152345 bp DNA linear VRT 04-OCT-2003
 LOCUS Zebrafish DNA sequence from clone CH211-129H14 in linkage group 17,
 DEFINITION complete sequence.
 ACCESSION BX005128
 VERSION BX005128.4 GI:37518236
 KEYWORDS HTG.
 SOURCE Danio rerio (zebrafish)
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 152345)

AUTHORS Sehra, H.
TITLE Direct Submission
JOURNAL Submitted (03-OCT-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
 On Oct 5, 2003 this sequence version replaced gi:30141705.

COMMENT

----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: zfish-help@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; SW: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep Repeat names beginning 'dr' were identified by the Recon repeat discovery system (Zhifeng Bao and Sean Eddy, submitted), and those beginning 'dtr' were identified by Rick Waterman (Stephen Johnson lab, WashU). For further information see

http://www.sanger.ac.uk/Projects/D_reiro/fishmask.shtml
 CH211-129H14 is from a CHORI-211 BAC library

VECTOR: PTARBAC2.1
 Clone-derived Zebrafish pUC subclones occasionally display inconsistency over the length of mononucleotide A/T runs and conserved TA repeats. Where this is found the longest good quality representation will be submitted.

FEATURES**source**

1.152345
 /organism="Danio rerio"
 /mol_type="genomic DNA"
 /db_xref="taxon:7955"
 /clone="CH211-129H14"
 /clone_lib="CHORI-211"

ORIGIN

Query Match 0.7%; Score 23; DB 5; Length 152345;
 Best Local Similarity 100.0%; Pred. No. 1.9;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3141 TTACTATTTCCTTTTCTTTTCTT 3163
 |||||
 DB 2054 TTAGTATTTCCTTTTCTTTTCTT 2032

RESULT 15**AC048381**

LOCUS AC048381 180925 bp DNA linear HTG 24-AUG-2002
 DEFINITION Homo sapiens chromosome 3 clone RP11-547K2 map 3, WORKING DRAFT
 AC048381
 SEQUENCE 37 unordered pieces.

ACCESSION AC048381.3 GI:8084282
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS

SOURCE Homo sapiens
 ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 180925)

AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
TITLE Homo sapiens chromosome 3, clone RP11-547K2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 180925)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Baettien, V., Beda, F., Bogunlavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Hayes, B., Heatford, A., Horton, L., Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lebecky, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McEwan, K., McPherson, R., Meldrum, J., Menus, L., Miho, T., Miranda, C., Miengo, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

TITLE Submitted (14-APR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL 3 (bases 1 to 180925)

REFERENCE**AUTHORS**

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Baettien, V., Beda, F., Bogunlavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Hayes, B., Heatford, A., Horton, L., Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lebecky, J., Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Margulis, N., McCarthy, M., McEwan, P., McGurk, A., McEwan, K., McPherson, R., Meldrum, J., Menus, L., Miho, T., Miranda, C., Miengo, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL On May 25, 2000 this sequence version replaced gi:770638.

COMMENT

All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/MIT Center for Genome Research
 Center code: WIR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
 Center project name: L9685
 Center clone name: 547_K_2

----- Summary Statistics
 Sequencing vector: M13; M7815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731

Consensus quality: 158029 bases at least Q40
 Consensus quality: 168695 bases at least Q30
 Consensus quality: 173285 bases at least Q20

Insert size: 183000; agarose-gel

Insert size: 177325; sum-of-ctrls
Quality coverage: 3.4 in Q20 bases; agarose-fp
Quality coverage: 3.6 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently consists of 37 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

1 1010: contig of 1010 bp in length
1011 1110: gap of 100 bp
1111 2487: contig of 1377 bp in length
2488 2587: gap of 100 bp
2588 3720: contig of 1133 bp in length
3721 3820: gap of 100 bp
3821 4892: contig of 1072 bp in length
4893 4992: gap of 100 bp
4993 5489: contig of 497 bp in length
5490 5589: gap of 100 bp
5590 7151: contig of 1562 bp in length
7152 7251: gap of 100 bp
7252 8655: contig of 1404 bp in length
8656 8755: gap of 100 bp
8756 10134: contig of 1379 bp in length
10135 10235: gap of 100 bp
10236 13107: contig of 2873 bp in length
13108 13207: gap of 100 bp
13208 15265: contig of 2058 bp in length
15266 15365: gap of 100 bp
15366 17768: contig of 2403 bp in length
17769 17868: gap of 100 bp
17869 20233: contig of 2365 bp in length
20234 20333: gap of 100 bp
20334 22968: contig of 2635 bp in length
22969 23068: gap of 100 bp
23069 25333: contig of 2265 bp in length
25334 25433: gap of 100 bp
25434 28978: contig of 3545 bp in length
28979 29078: gap of 100 bp
29079 31604: contig of 2526 bp in length
31605 31704: gap of 100 bp
31705 34400: contig of 2696 bp in length
34401 34501: gap of 100 bp
34502 37255: contig of 2755 bp in length
37256 37355: gap of 100 bp
37356 40057: contig of 2702 bp in length
40058 40157: gap of 100 bp
40158 44190: contig of 4033 bp in length
44191 44290: gap of 100 bp
44291 46785: contig of 2495 bp in length
46786 46885: gap of 100 bp
46886 52597: contig of 5712 bp in length
52598 52697: gap of 100 bp
52698 59026: contig of 6329 bp in length
59027 59126: gap of 100 bp
59127 64431: contig of 5305 bp in length
64432 64531: gap of 100 bp
64532 69115: contig of 4584 bp in length
69116 69215: gap of 100 bp
69216 73865: contig of 4651 bp in length
73867 73965: gap of 100 bp
73967 79516: contig of 5550 bp in length
79517 79616: gap of 100 bp
79617 85688: contig of 6072 bp in length
85689 85789: gap of 100 bp
85790 92058: contig of 6270 bp in length
92059 92158: gap of 100 bp
92159 99039: contig of 6881 bp in length
99040 99139: gap of 100 bp
108885: contig of 9746 bp in length

```

```

* 108886 108985: gap of 100 bp
* 108986 116971: contig of 7986 bp in length
* 116972 117071: gap of 100 bp
* 117072 126718: contig of 9647 bp in length
* 126719 126818: gap of 100 bp
* 126819 137354: contig of 10536 bp in length
* 137355 137455: gap of 100 bp
* 137455 146943: contig of 9489 bp in length
* 146944 147043: gap of 100 bp
* 147044 157785: contig of 10742 bp in length
* 157786 157885: gap of 100 bp
* 157886 180925: contig of 23040 bp in length.
* Location/Qualifiers
  1. 180925
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="3"
    /map="3"
    /clone="RP11-547K2"
    /clone_11b="RP11 Human Male BAC"
    1. 1010
      /note="assembly_fragment"
    1111. 2487
      /note="assembly_fragment"
    2588. 3720
      /note="assembly_fragment"
    3821. 4892
      /note="assembly_fragment"
    4993. 5489
      /note="assembly_fragment"
      clone_end:T7
      vector_side:right"
    5590. 7151
      /note="assembly_fragment"
    7252. 8655
      /note="assembly_fragment"
    8756. 10134
      /note="assembly_fragment"
    10235. 13107
      /note="assembly_fragment"
    13208. 15265
      /note="assembly_fragment"

```

```

Query Match 0.7%; Score 23; DB 2; Length 180925;
Best Local Similarity 100.0%; Pred.No.1.9;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1673 TCTTTACTCTTAATTTTCT 1695
|||
Db 16692 TCTTTACTCTTAATTTTCT 16714

```

Search completed: March 30, 2004, 14:48:22
Job time: 12264.4 secs

This Page Blank (uspto)